(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AATGCGGCCA CTCCAAGGAG GCCGGGAGGA TTGTGGGAGG CCAAGACACC CAGGAAGGAC 60

GCTGGCCGTG GCAGGTTGGC CTGTGGTTGA CCTCAGTGGG GCATGTATGT GGGGGCTCCC 120

TCATCCACCC ACGCTGGGTG CTCACAGCCG CCCACTGCTT CCTGAGGTCT GAGGATCCCG 180

GGCTCTACCA TGTTAAAGTC GGAGGGCTGA CACCCTCACT TTCAGAGCCC CACTCGGCCT 240

TGGTGGCTGT GAGGAGGCTC CTGGTCCACT CCTCATACCA TGGGACCACC ACCAGCGGGG 300

ACATTGCCCT GATGGAGCTG GACTCCCCCT TGCAGGCCTC CCAGTTCAGC CCCATCTGCC 360

TCCCAGGACC CCAGACCCCC CTCGCCATTG GGACCGTGTG CTGGGTAAAC GGGCTGGGGG 420

TCCACTCAGG AGAGGCCCTG GCGAGTGTCC TTCAGGAGGT GGCTGTGCCC CTCCTGGACT 480

CGAACATGTG TGAGCTGATG TACCACCTAG GAGAGCCCAG CCTGGCTGGC CAGCGCCTCA 540

TCCAGGACGA CATGCTCTGT GCTGGCTCTG TCCAGGGCAA GAAAGACTCC TGCCAGGGTG 600

ACTCCGGGGG GCCGCTGGTC TGCCCCATCA ATGATACGTG GATCCAGGCC GGCATTGTGA 660

GCTGGGGATT CGGCTGTGCC CGGCCTTTCC GGCCTGGTGT CTACACCCAG GTGCTAAGCT 720

ACACAGACTG GATTCAGAGA ACCCTGGCTG AATCTCACTC AGGCATGTCT GGGGCCCGCC 780

CAGGTGCCCC AGGATCCCAC TCAGGCACCT CCAGATCCCA CCCAGTGCTG CTGCTTGAGC 840

TGTTGACCGT ATGCTTGCTT GGGTCCCTGT GAACCATGAG CCATGGAGTC CGGGATCCCC 900

TTTCTGGTAG GATTGATGGA ATCTAATAAT AAA 933

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 980 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
CCTGTGGTCG CCCCAGGATG CTGAACCGAA TGGTGGGCGG GCAGGACACG CAGGAGGGCG      60
AGTGGCCCTG GCAAGTCAGC ATCCAGCGCA ACGGAAGCCA CTTCTGCGGG GGCAGCCTCA     120
TCGCGGAGCA GTGGGTCCTG ACGGCTGCGC ACTGCTTCCG CAACACCTCT GAGACGTCCC     180
TGTACCAGGT CCTGCTGGGG GCAAGGCAGC TAGTGCAGCC GGGACCACAC GCTATGTATG     240
CCCGGGTGAG GCAGGTGGAG AGCAACCCCC TGTACCAGGG CACGGCCTCC AGCGCTGACG     300
TGGCCCTGGT GGAGCTGGAG GCACCAGTGC CCTTCACCAA TTACATCCTC CCCGTGTGCC     360
TGCCTGACCC CTCGGTGATC TTTGAGACGG GCATGAACTG CTGGGTCACT GGCTGGGGCA     420
GCCCCAGTGA GGAAGACCTC CTGCCCGAAC CGCGGATCCT GCAGAAACTC GCTGTGCCCA     480
TCATCGACAC ACCCAAGTGC AACCTGCTCT ACAGCAAAGA CACCGAGTTT GGCTACCAAC     540
CCAAAACCAT CAAGAATGAC ATGCTGTGCG CCGGCTTCGA GGAGGGCAAG AAGGATGCCT     600
GCAAGGGCGA CTCGGGCGGC CCCCTGGTGT GCCTCGTGGG TCAGTCGTGG CTGCAGGCGG     660
GGGTGATCAG CTGGGGTGAG GGCTGTGCCC GCCAGAACCG CCCAGGTGTC TACATCCGTG     720
TCACCGCCCA CCACAACTGG ATCCATCGGA TCATCCCCAA ACTGCAGTTC CAGCCAGCGA     780
GGTTGGGCGG CCAGAAGTGA GACCCCCGGG GCCAGGAGCC CCTTGAGCAG AGCTCTGCAC     840
CCAGCCTGCC CGCCCACACC ATCCTGCTGG TCCTCCCAGC GCTGCTGTTG CACCTGTGAG     900
CCCCACCAGA CTCATTTGTA AATAGCGCTC CTTCCTCCCC TCTCAAATAC CCTTATTTTA     960
TTTATGTTTC TCCCAATAAA                                                 980
```

CLAIMS:

1. An isolated proteinaceous molecule involved in or associated with regulation of cell activity and/or viability comprising a sequence of amino acids encoded by a nucleotide sequence, at least a portion of which, is capable of being amplified by polymerase chain reaction (PCR) using the following primers:

5′ ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3′ [SEQ ID NO:1]; and

5′ACAGAATTCAXIGGICCICCIC/GT/AXTCICC3′ [SEQ ID NO:2];

or a complementary form of said primers.

2. An isolated proteinaceous molecule according to claim 1 wherein said molecule is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:4 or an amino acid sequence having at least 50% similarity thereto.

3. An isolated proteinaceous molecule according to claim 1 wherein said molecule is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least 50% similarity thereto.

4. An isolated proteinase molecule according to claim 1 wherein said molecule is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity thereto.

5. An isolated proteinaceous molecule according to claim 1 wherein said molecule is a serine proteinase comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions at 42°C.

6. An isolated proteinaceous molecule according to claim 1 wherein said molecule is a serine proteinase comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:5 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions at 42°C.

7. An isolated proteinaceous molecule according to claim 1 wherein said molecule is a serine proteinase comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions at 42°C.

8. An isolated proteinaceous molecule according to claim 1 wherein said molecule is a kinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:10 or having 50% amino acid similarity thereto.

9. An isolated proteinaceous molecule according to claim 1 wherein said molecule is a kinase comprising an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:9 under low stringency conditions at 42°C.

10. An isolated nucleic acid molecule encoding a polypeptide wherein at least a portion of said nucleic acid molecule is capable of being amplified by polymerase chain reaction (PCR) using the following primers:

5′ ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3′ [SEQ ID NO:1]; and

5′ACAGAATTCAXIGGICCICCIC/GT/AXTCICC3′ [SEQ ID NO:2];

or a complementary form of said primers.

11. An isolated nucleic acid molecule according to claim 10 wherein said polypeptide is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:4 or an amino acid sequence having at least 50% similarity thereto.

12. An isolated nucleic acid molecule according to claim 10 wherein said polypeptide is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least 50% similarity thereto.

13. An isolated nucleic acid molecule according to claim 10 wherein said polypeptide is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity thereto.

14. An isolated nucleic acid molecule according to claim 10 comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:3 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions at 42°C.

15. An isolated nucleic acid molecule according to claim 10 comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:5 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions at 42°C.

16. An isolated nucleic acid molecule according to claim 10 comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions at 42°C.

17. An isolated nucleic acid molecule according to claim 10 wherein said polypeptide is a kinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:10 or having 50% amino acid similarity thereto.

18. An isolated nucleic acid molecule according to claim 17 comprising a sequence of nucleotides encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:9 under low stringency conditions at 42°C.

19. An isolated serine proteinase encoded by a gene proximal to a cluster of genes of a mammalian chromosome.

20. An isolated serine proteinase according to claim 19 wherein the mammalian chromosome is human chromosome 16p13.3 or its equivalent in a non-human species.

21. An isolated serine proteinase according to claim 20 wherein the gene cluster includes at least two genes having the nucleotide sequence as set forth in SEQ ID NO:3 or 5 or 28 or 29 or 30 or a nucleotide sequence having at least 50% similarity to any one of SEQ ID NO:3 or 5 or 28 or 29 or 30 or a nucleotide sequence capable of hybridizing to any one of the sequences under low stringency conditions at 42°C.

22. An isolated serine proteinase according to claim 20 wherein said serine proteinase is a short form of HELA2 having an amino acid sequence substantially as set forth in SEQ ID NO:4 or an amino acid sequence having at least 50% similarity thereto.

23. An isolated serine proteinase according to claim 20 wherein said serine proteinase is a long form of HELA2 having an amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least 50% similarity thereto.

24. An isolated serine proteinase according to claim 22 encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or a nucleotide sequence having at least 50% similarity thereto or a sequence capable of hybridizing to SEQ I NO:3 under low stringency conditions at 42°C.

25. An isolated serine proteinase according to claim 23 encoded by a nucleotide sequence substantially as set forth in SEQ ID NO;5 or a nucleotide sequence having at least 50% similarity thereto or a sequence capable of hybridizing to SEQ ID NO:5 under low stringency conditions at 42°C.

26. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a serine proteinase and corresponding to a gene proximal to a cluster of genes encoding serine proteinases.

27. An isolated nucleic acid molecule according to claim 26 wherein the gene cluster includes at least two genes having the nucleotide sequence as set forth in SEQ ID NO:3 or 5 or 28 or 29 or 30 or a nucleotide sequence having at least 50% similarity to any one of SEQ ID NO:3 or 5 or 28 or 29 or 30 or a nucleotide sequence capable of hybridizing to any one of the sequences under low stringency conditions at 42°C.

28. An isolated nucleic acid molecule according to claim 25 comprising a nucleotide sequence substantially as set forth in SEQ ID NO:3 or SEQ ID NO:5 or a nucleotide sequence having at least about 50% similarity to either of SEQ ID NO:3 or SEQ ID NO:5 or a nucleotide sequence capable of hybridizing to SEQ ID NO:3 or SEQ ID NO:5 under low stringency conditions at 42°C.

29. An isolated kinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:10 or an amino acid sequence having at least about 50% similarity thereto.

30. An isolated kinase according to claim 29 encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity thereto or capable of hybridizing to SEQ ID NO:9 under low stringency conditions at 42°C.

31. A method of regulating cell activity and/or viability said method comprising contacting said cell with an activity and/or viability effective amount of a serine proteinase and/or kinase.

32. A method according to claim 31 wherein the serine proteinase comprises a sequence of amino acids encoded by a nucleotide sequence, at least a portion of which, is capable of being amplified by polymerase chain reaction (PCR) using the following primers:

5′ ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3′ [SEQ ID NO:1]; and

5′ACAGAATTCAXIGGICCICCIC/GT/AXTCICC3′ [SEQ ID NO:2];

or a complementary form of said primers.

33. A method according to claim 31 wherein the serine proteinase comprises an amino acid sequence substantially as set forth in SEQ ID NO:4 or an amino acid sequence having at least 50% similarity thereto.

34. A method according to claim 31 wherein the serine proteinase comprises an amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least 50% similarity thereto.

35. A method according to claim 31 wherein the serine proteinase comprises an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity thereto.

36. A method according to claim 31 wherein the serine proteinase comprises a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions at 42°C.

37. A method according to claim 31 wherein the serine proteinase comprises a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:5 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions at 42°C.

38. A method according to claim 31 wherein the serine proteinase comprises a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions at 42°C.

39. A method according to claim 31 wherein the kinase comprises an amino acid sequence substantially as set forth in SEQ ID NO:10 or having 50% amino acid similarity thereto.

40. A method according to claim 31 wherein the kinase comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:9 under low stringency conditions at 42°C.

41. A method of modulating fertility in a mammal said method comprising modulating levels of HELA2 wherein increasing levels of HELA2 facilitates sperm maturation and development.

42. A method according to claim 41 wherein fertility is enhanced by introducing recombinant HELA2.

43. A method according to claim 41 wherein fertility is reduced by down regulating expression of the HELA2 gene.

44. A composition comprising a serine proteinase and/or kinase capable of regulating cell activity and/or viability and one or more pharmaceutically acceptable carriers and/or diluents.

45. A composition according to claim 44 wherein the serine proteinase is HELA2 or a functional derivative thereof.

46. An isolated antibody capable of interacting with a proteinaceous molecule involved in or associated with regulation of cell activity and/or viability comprising a sequence of amino acids encoded by a nucleotide sequence, at least a portion of which, is capable of being

amplified by polymerase chain reaction (PCR) using the following primers:

5′ ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3′ [SEQ ID NO:1]; and

5′ACAGAATTCAXIGGICCICCIC/GT/ÀXTCICC3′ [SEQ ID NO:2];

or a complementary form of said primers.

47. An isolated antibody according to claim 46 wherein said proteinaceous molecule is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:4 or an amino acid sequence having at least 50% similarity thereto.

48. An isolated antibody according to claim 46 wherein said proteinaceous molecule is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least 50% similarity thereto.

49. An isolated antibody according to claim 46 wherein said proteinaceous molecule is a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity thereto.

50. An isolated antibody according to claim 46 wherein said proteinaceous molecule is a serine proteinase comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions at 42°C.

51. An isolated antibody according to claim 46 wherein said proteinaceous molecule is a serine proteinase comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:5 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions at 42°C.

52. An isolated antibody according to claim 46 wherein said proteinaceous said molecule is a serine proteinase comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions at 42°C.

53. An isolated antibody according to claim 46 wherein said proteinaceous molecule is a kinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:10 or having 50% amino acid similarity thereto.

54. An isolated antibody according to claim 46 wherein said proteinaceous molecule is a kinase comprising an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:9 under low stringency conditions at 42°C.

55. An antagonist or agonist to the isolated proteinaceous molecule according to any one of claims 1 to 9.

56. A method of determining a predisposition for or the presence of a cancer, said method comprising determining the presence of a nucleotide sequence encoding a proteinaceous molecule according to any one of claims 1 to 9.

57. A method according to claim 56 wherein the nucleotide sequence encodes a polypeptide wherein at least a portion of said nucleotide sequence is capable of being amplified by polymerase chain reaction (PCR) using the following primers:

5′ ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3′ [SEQ ID NO:1]; and

5′ACAGAATTCAXIGGICCICCIC/GT/AXTCICC3′ [SEQ ID NO:2];

or a complementary form of said primers.

58. A method according to claim 57 wherein said nucleotide sequence encodes a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:4 or an amino acid sequence having at least 50% similarity thereto.

59. A method according to claim 57 wherein said nucleotide sequence encodes a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least 50% similarity thereto.

60. A method according to claim 57 wherein said nucleotide sequence encodes a serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity thereto.

61. A method according to claim 57 wherein said nucleotide sequence is as substantially set forth in SEQ ID NO:3 or is a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions at 42°C.

62. A method according to claim 57 wherein said nucleotide sequence is as substantially set forth in SEQ ID NO:5 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions at 42°C.

63. A method according to claim 57 wherein said nucleotide sequence is as substantially set forth in SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions at 42°C.

64. A method according to claim 57 wherein said nucleotide sequence is as substantially set forth in SEQ ID NO:10 or having 50% amino acid similarity thereto.

65. A method according to claim 57 wherein said nucleotide sequence is as substantially set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity thereto or a nucleotide sequence capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:9 under low stringency conditions at 42°C.

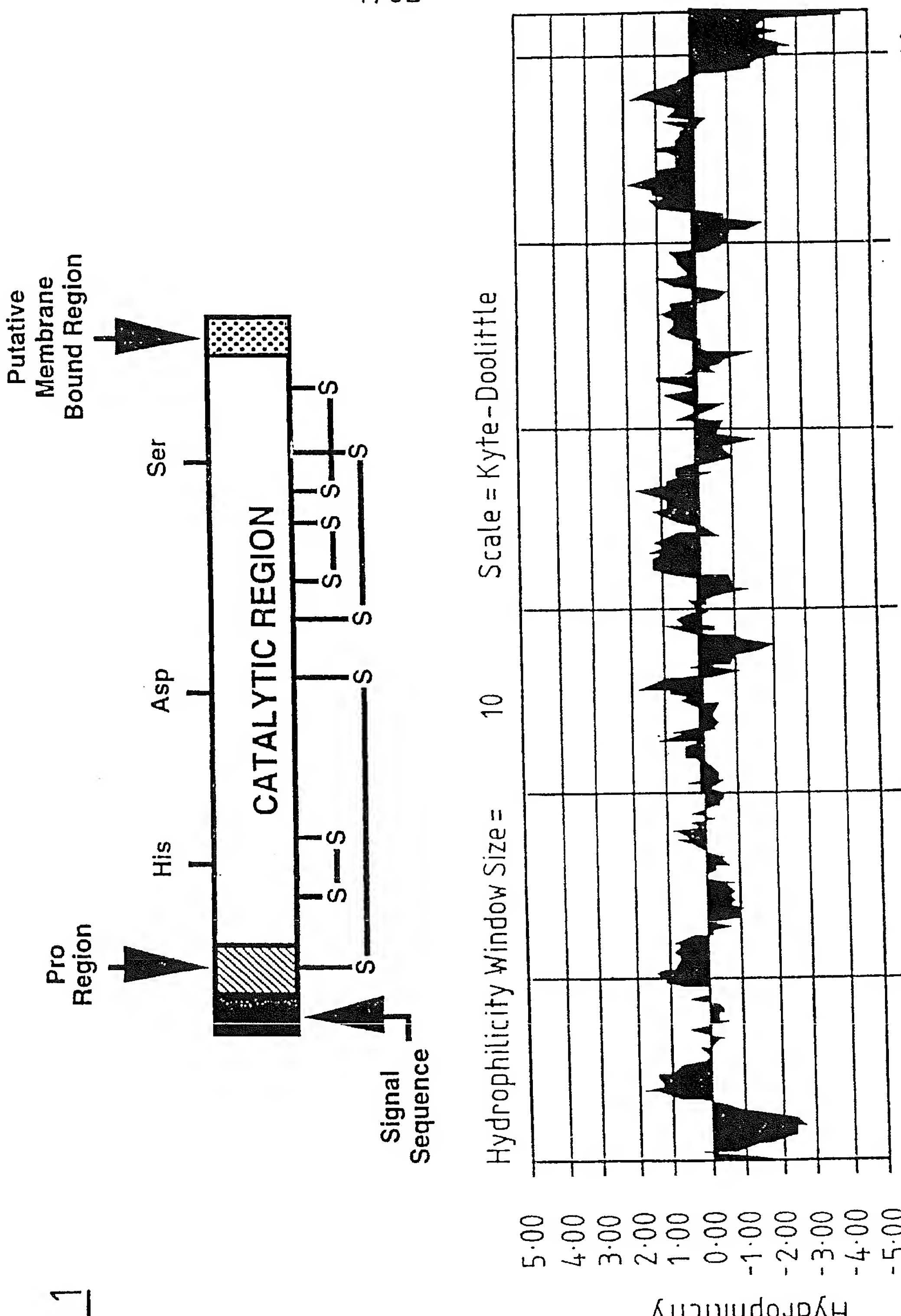
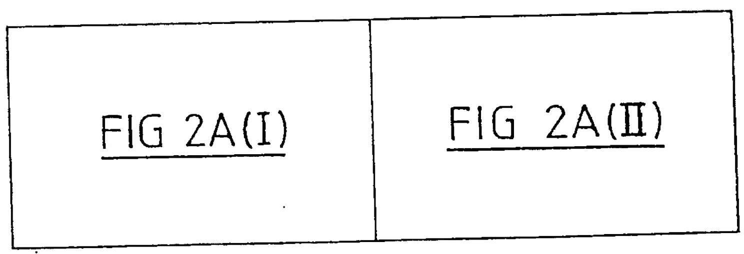

Putative Membrane Bound Region
Ser
Asp
His
Pro Region
CATALYTIC REGION
S—S
Signal Sequence
Hydrophilicity Window Size = 10 Scale = Kyte-Doolittle
50
100
150
200
250
300
5·00
4·00
3·00
2·00
1·00
0·00
-1·00
-2·00
-3·00
-4·00
-5·00
Hydrophilicity

FIG 1

FIG 2A



Sequence comparison of HELA2(Testisin) and prostasin

```
                Signal sequence               Light Chain
prostasin  MAQKGVLGPGQLGAVAILLYLGLLRSGTGAEGA--EAPCG-VAPQARITGGSSAVA
HELA2      MGARGAL----L--LALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAEL
           * .* *   *  .*.**   **    * *   ***     .** **  *
                                             S-S
                              ASP                           N-gly
                               *
prostasin  KVSTLKDIIPHPSYLQEGSQGDIALLQLSRPITFSRYIRPICLPAANASFPNGLHC
HELA2      TRYFVSNIYLSPRYLG-NSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDC
            . *  * **  .*  ****..** *.*... *.**** *..  * *   *
                                                         N-gly
                                     S-S
      SER
       *
prostasin  ACQGDSGGPLSCPVEGLWYLTGIVSWGDACGARNRPGVYTLASSYASWIQSKVTEL
HELA2      ACFGDSGGPLACNKDGLWYQIGVVSWGVGCGRPNRPGVYTNISHHFEWIQ------
           ** *******.*  .****  *.****  **  *******  *    ***
                        S-S                    N-gly
```

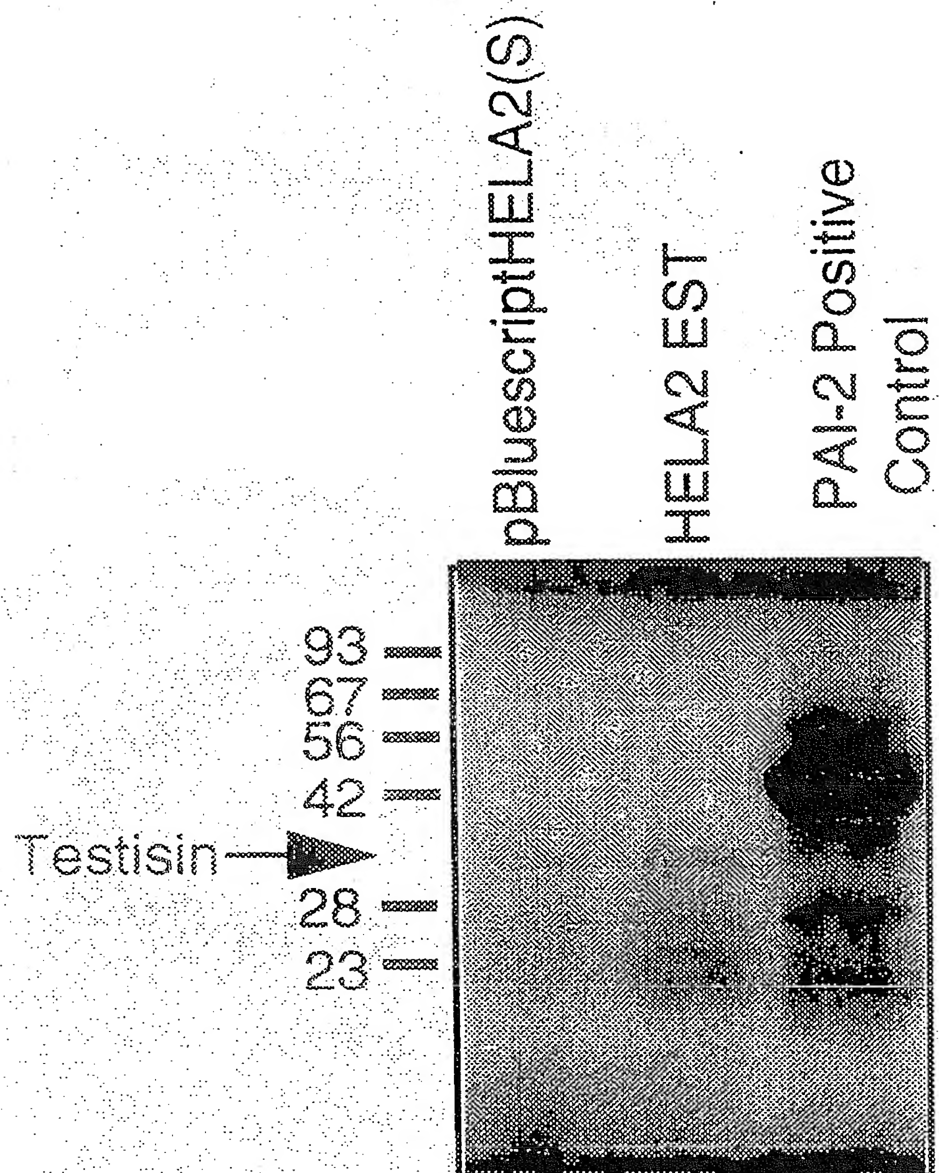
FIG 2A(I)

Sequence comparison of HELA2(Testisin) and prostasin

```
                                                HIS
                                                 *
prostasin   GQWPWQVSITYEGVHVCGGSLVSEQWVLSAAHCFPSEHHK-EAYEVKLG----AHQLDSYSEDA
HELA2       GRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETDLSDPSGWMVQFGQLTSMPSFWSLQAYY
            *.**** *.       **** **.* .* *.*****  ..      *.*        *
                            |_____S-S_____|         YS
                                              (Long Isoform)

                              Heavy Chain
prostasin   TVTGWGHVAPSVSLLTPKPLQQLEVPLISRETCNCLYNIDAKPEEPHFVQEDMVCAGYVEGGKD
HELA2       WVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKD--IFG-DMVCAGNAQGGKD
            *****  .    .* .*. **...* .* ** *.  .  .          ******  .****
                                     N-gly
                                       |_________S-S__________|

                                                   Putative
                                                   Transmembrane
                                                   Domain
prostasin   QPRVVPQTQESQPDSNLCGSHLAFSSAPAQGLLR     PILFLPLGLALGLLSPWL   SEH
HELA2       --KLMAQSGMSQPD-------------PS----W     PLLFFPLLWALPLLGPV
              ...*. ****               *.          *.** **  ** ** *
```

FIG 2A(II)

FIG 2B



In vitro transcription / translation of HELA2 (Testisin)

FIG 3

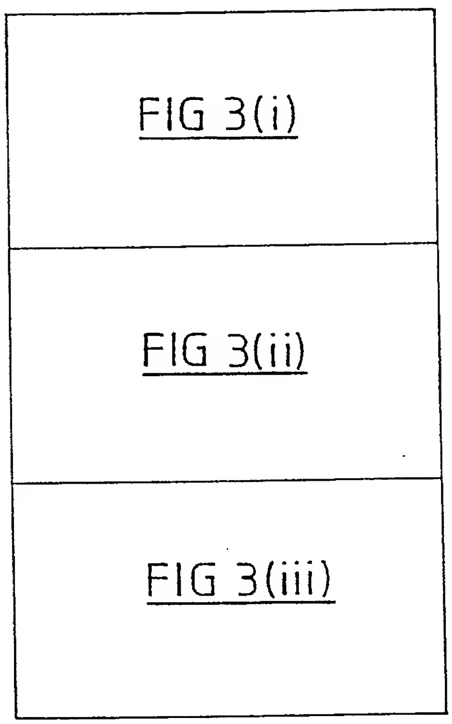
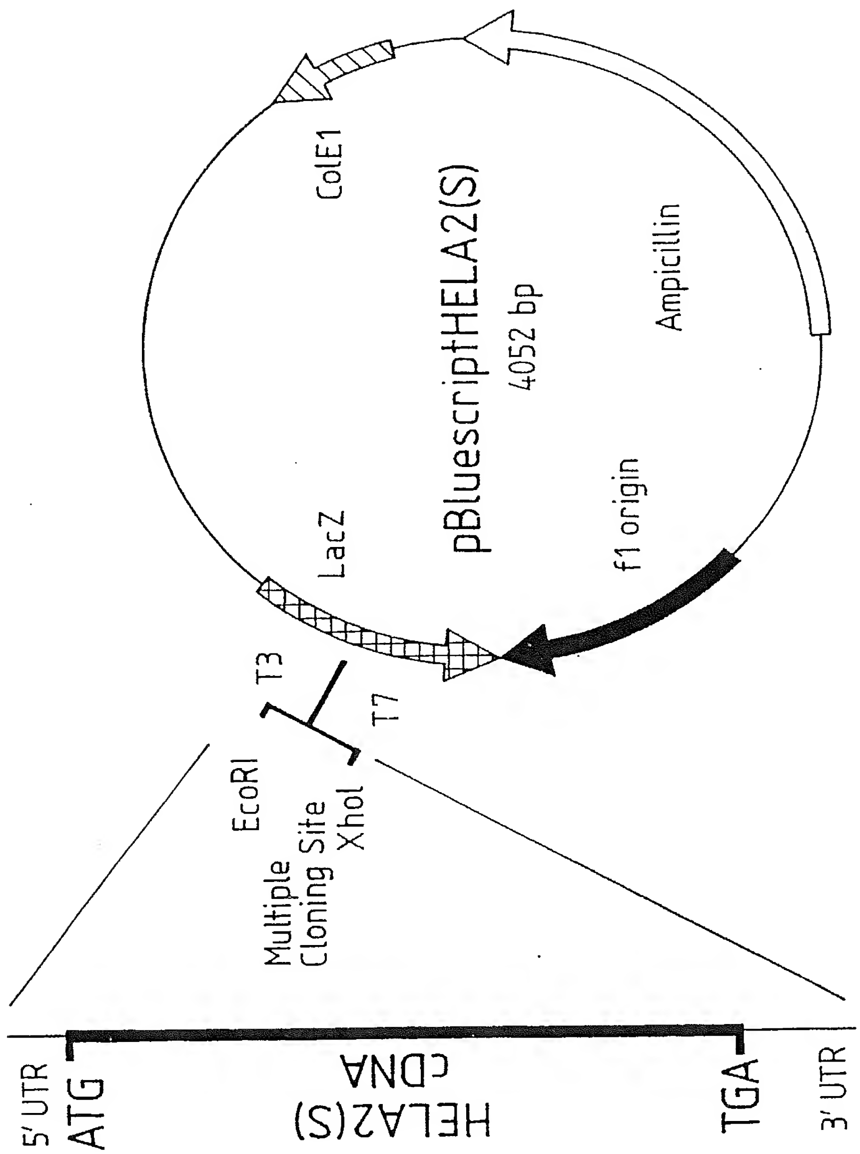
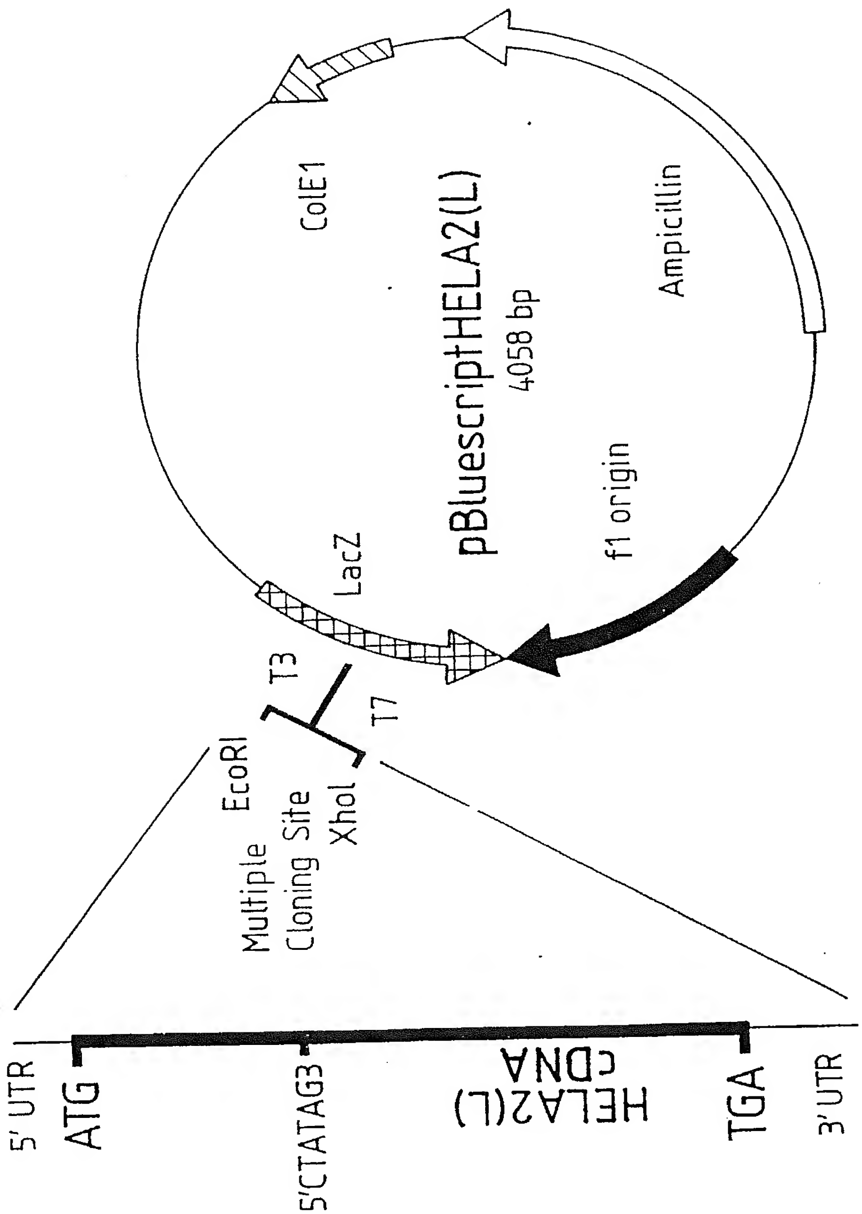


FIG 3(i)



HELA2 (Testisin) Short Isoform

FIG 3(ii)



HELA2 (Testisin) Long Isoform

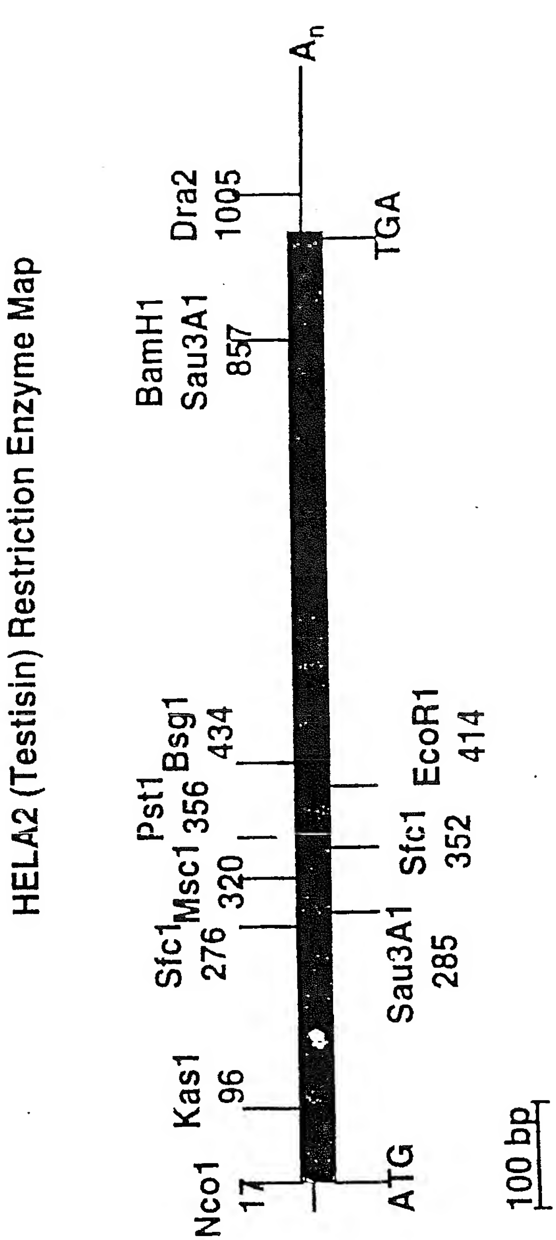
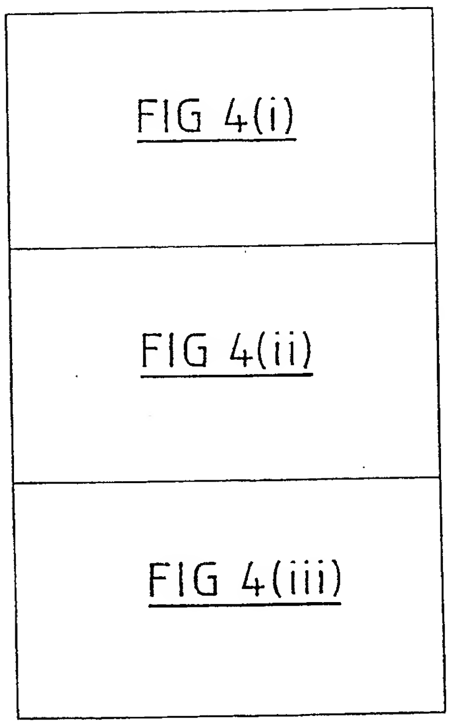

HELA2 (Testisin) Restriction Enzyme Map
Nco1 17
Kas1 96
Sfc1 276
Msc1 320
Pst1 356
Bsg1 434
BamH1 Sau3A1 857
Dra2 1005
A_n
ATG
TGA
Sau3A1 285
Sfc1 352
EcoR1 414
100 bp

FIG 3(iii)

FIG 4



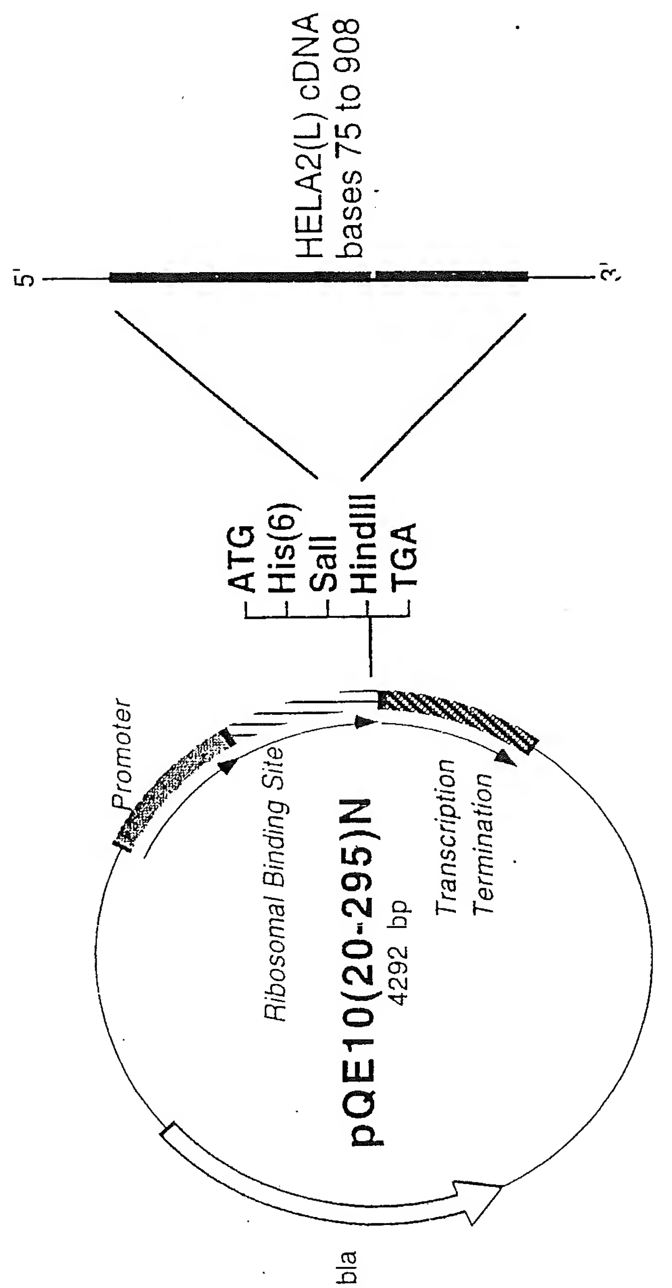

HELA2(L) cDNA
bases 75 to 908
5'
3'
ATG
His(6)
SalI
HindIII
TGA
Promoter
Ribosomal Binding Site
pQE10(20-295)N
4292 bp
Transcription
Termination
bla

FIG 4(i)

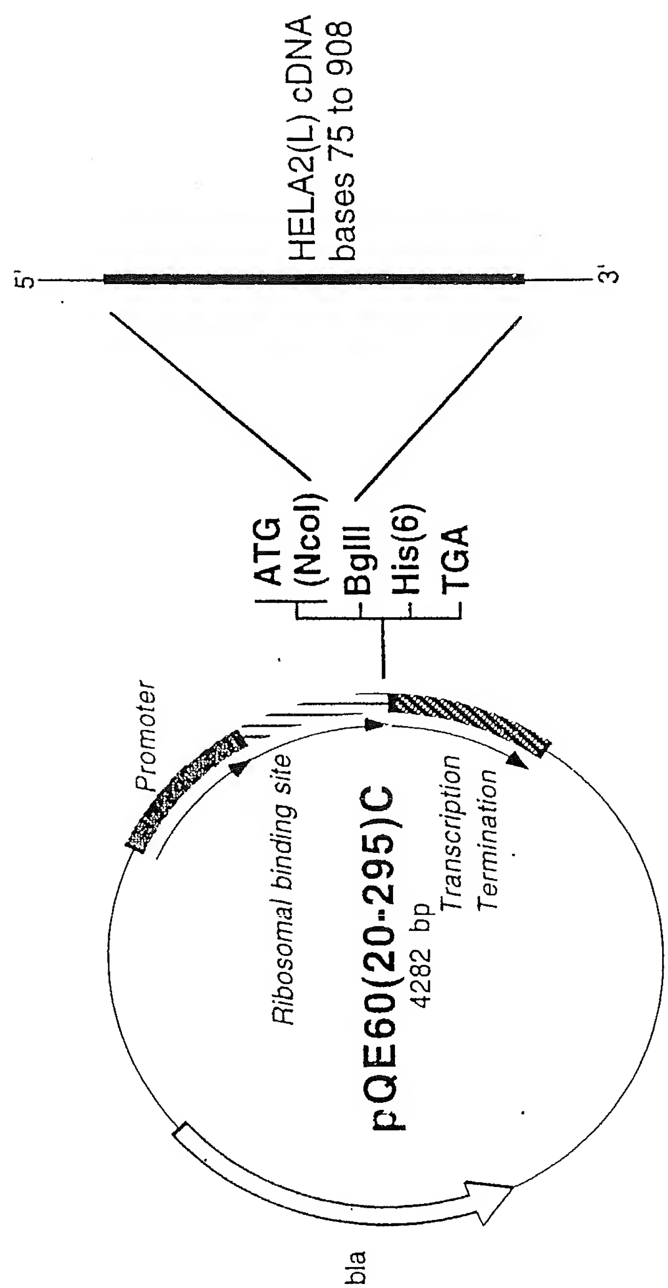
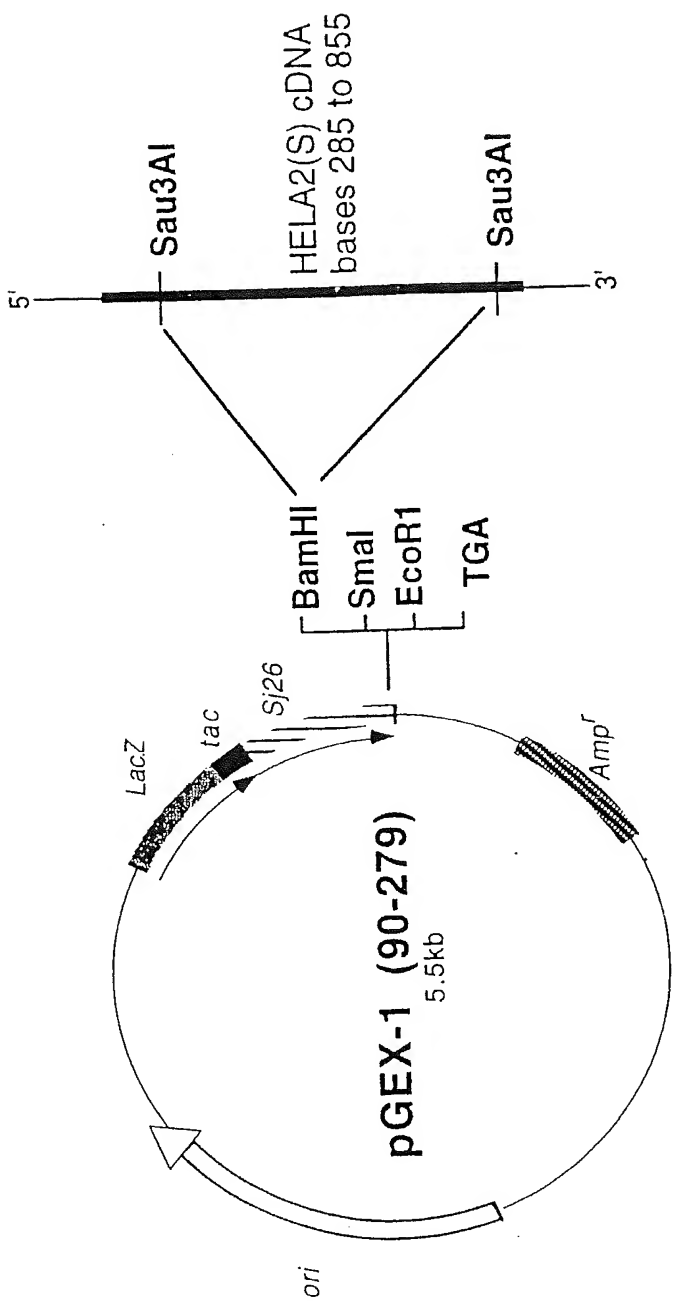

5'
3'
HELA2(L) cDNA
bases 75 to 908
ATG
(NcoI)
BglII
His(6)
TGA
Promoter
Ribosomal binding site
pQE60(20-295)C
4282 bp
Transcription
Termination
bla

FIG 4(ii)

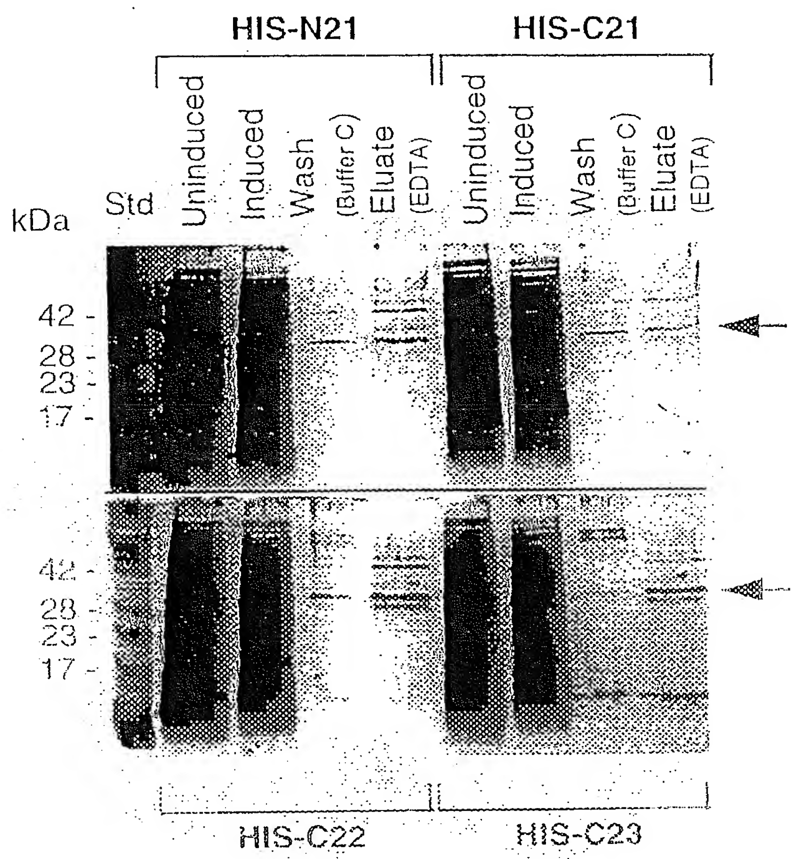


5'
3'
Sau3AI
HELA2(S) cDNA
bases 285 to 855
Sau3AI
BamHI
SmaI
EcoR1
TGA
Sj26
tac
LacZ
Amp^r
pGEX-1 (90-279)
5.5kb
ori

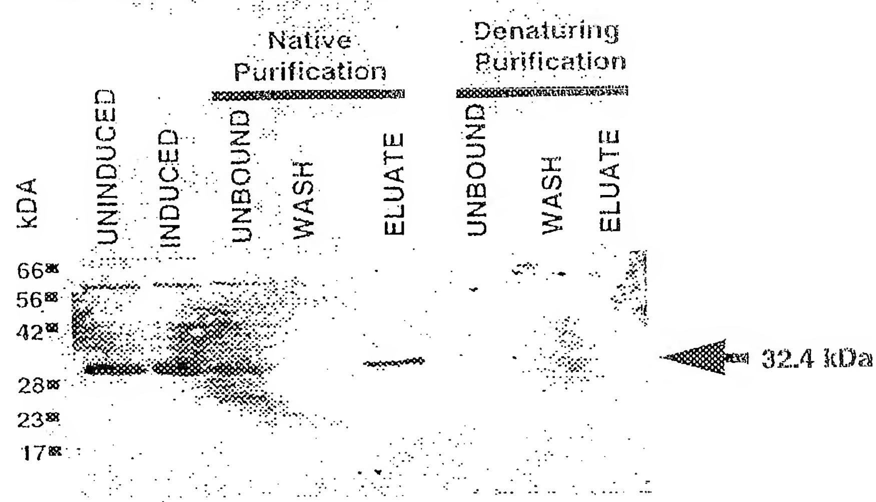
FIG 4(iii)

FIG 5

A. Expression of recombinant Testisin in *E. coli.*



B. Western blot of recombinant Testisin



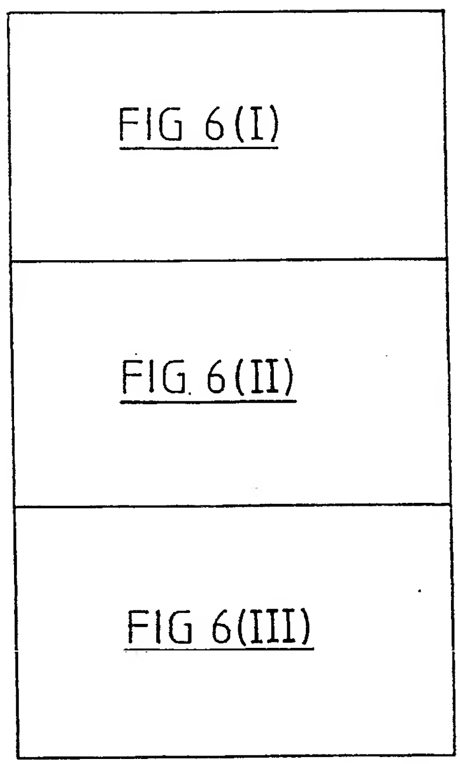


FIG 6

FIGURE 6(I)

```
1                                                   GCCGCGGGAGAGGAGGCC
19   ATGGGCGCGCGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCTGGACTCAGGAAG
      M  G  A  R  G  A  L  L  L  A  L  L  L  A  R  A  G  L  R  K   20

79   CCGGAGTCGCAGGAGGCGGCGCCGTTATCAGGACCATGCGGCCGACGGGTCATCACGTCG
      P  E  S  Q  E  A  A  P  L  S  G  P  C  G  R  R  V  I  T  S   40

139  CGCATCGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGC
      R  I  V  G  G  E  D  A  E  L  G  R  W  P  W  Q  G  S  L  R   60

199  CTGTGGGATTCCCACGTATGCGGAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCG
      L  W  D  S  H  V  C  G  V  S  L  L  S  H  R  W  A  L  T  A   80

259  GCGCACTGCTTTGAAACCTATAGTGACCTTAGTGATCCCTCCGGGTGGATGGTCCAGTTT
      A  H  C  F  E  T  Y  S  D  L  S  D  P  S  G  W  M  V  Q  F  100

319  GGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGTTACTTC
      G  Q  L  T  S  M  P  S  F  W  S  L  Q  A  Y  Y  T  R  Y  F  120

379  GTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTG
      V  S  N  I  Y  L  S  P  R  Y  L  G  N  S  P  Y  D  I  A  L  140
```

FIGURE 6 (II)

```
439  GTGAAGCTGTCTGCACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCC
      V  K  L  S  A  P  V  T  Y  T  K  H  I  Q  P  I  C  L  Q  A  160

499  TCCACATTTGAGTTTGAGAACCGGACAGACTGCTGGGTGACTGGCTGGGGGTACATCAAA
      S  T  F  E  F  E  N  R  T  D  C  W  V  T  G  W  G  Y  I  K  180

559  GAGGATGAGGCACTGCCATCTCCCCACACCCTCCAGGAAGTTCAGGTCGCCATCATAAAC
      E  D  E  A  L  P  S  P  H  T  L  Q  E  V  Q  V  A  I  I  N  200

619  AACTCTATGTGCAACCACCTCTTCCTCAAGTACAGTTTCCGCAAGGACATCTTTGGAGAC
      N  S  M  C  N  H  L  F  L  K  Y  S  F  R  K  D  I  F  G  D  220

679  ATGGTTTGTGCTGGCAATGCCCAAGGCGGGAAGGATGCCTGCTTCGGTGACTCAGGTGGA
      M  V  C  A  G  N  A  Q  G  G  K  D  A  C  F  G  D  S  G  G  240

739  CCCTTGGCCTGTAACAAGAATGGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTG
      P  L  A  C  N  K  N  G  L  W  Y  Q  I  G  V  V  S  W  G  V  260

799  GGCTGTGGTCGGCCCAATCGGCCCGGTGTCTACACCAATATCAGCCACCACTTTGAGTGG
      G  C  G  R  P  N  R  P  G  V  Y  T  N  I  S  H  H  F  E  W  280
```

FIGURE 6 (III)

```
859   ATCCAGAAGCTGATGGCCCAGAGTGGCATGTCCCAGCCAGACCCCTCCTGGCCGCTACTC
        I  Q  K  L  M  A  Q  S  G  M  S  Q  P  D  P  S  W  P  L  L  300

919   TTTTTCCCTCTTCTCTGGGCTCTCCCACTCCTGGGGCCGGTCTGAGCCTACCTGAGCCCA 314
        F  F  P  L  L  W  A  L  P  L  L  G  P  V  *

979   TGCAGCCTGGGGCCACTGCCAAGTCAGGCCCTGGTTCTCTTCTGTCTTGTTTGGTAATAA
1039  ACACATTCCAGTTGATGCCTTGCAGGGCATTCTTCAAAAAAAAAAAAAAAAAAAAAAAAA
1099  AAAAAAAAAAAAAAAAAAA
```

Western blot of GST-Testisin using anti-Testisin peptide T175 antibody

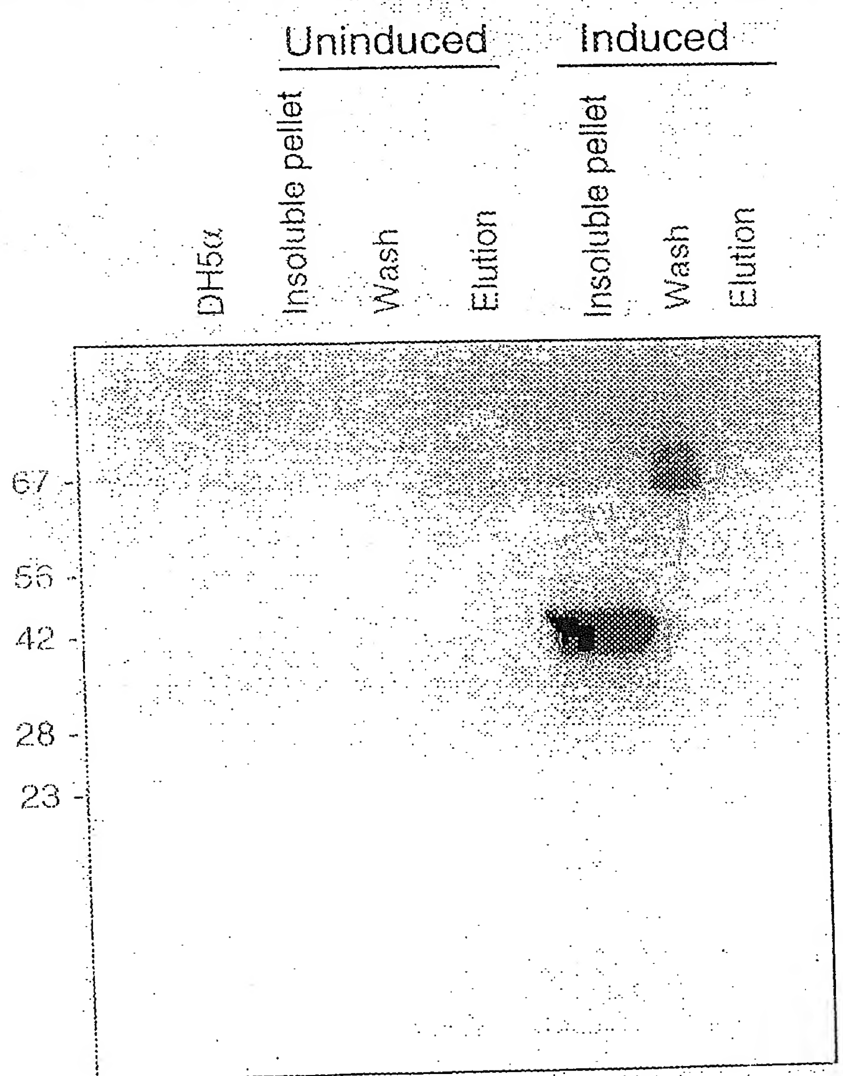
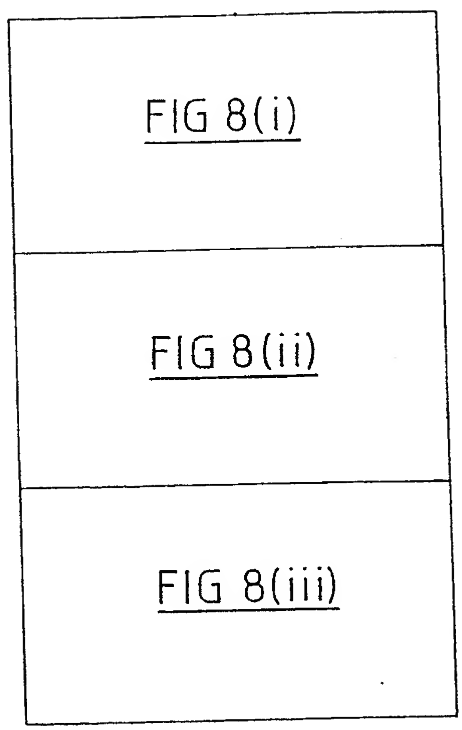


FIG 7

FIG 8



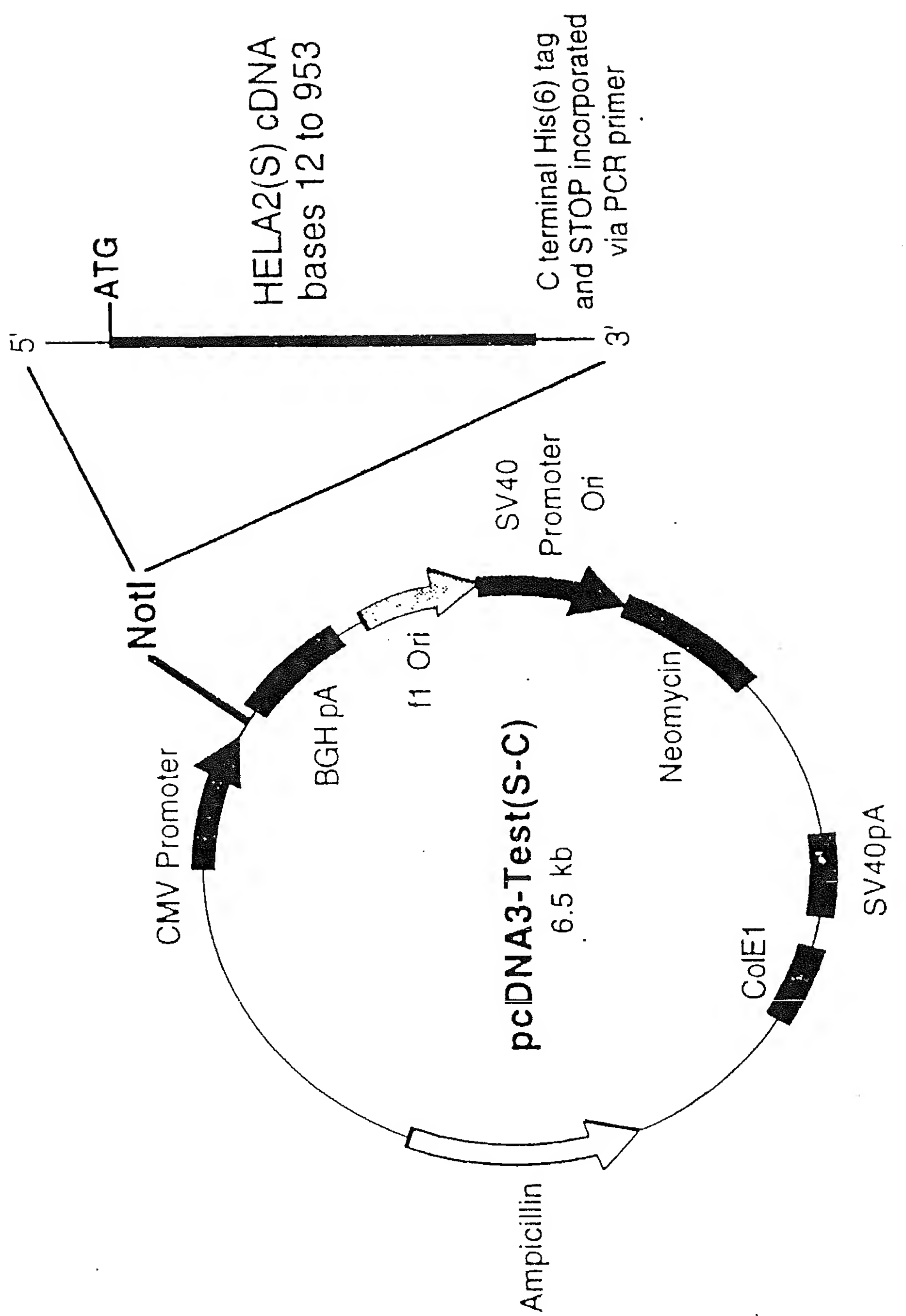

HELA2(S) cDNA
bases 12 to 953
ATG
5'
3'
C terminal His(6) tag
and STOP incorporated
via PCR primer
NotI
SV40
Promoter
Ori
CMV Promoter
BGH pA
f1 Ori
Neomycin
pcDNA3-Test(S-C)
6.5 kb
SV40pA
ColE1
Ampicillin

FIG 8(i)

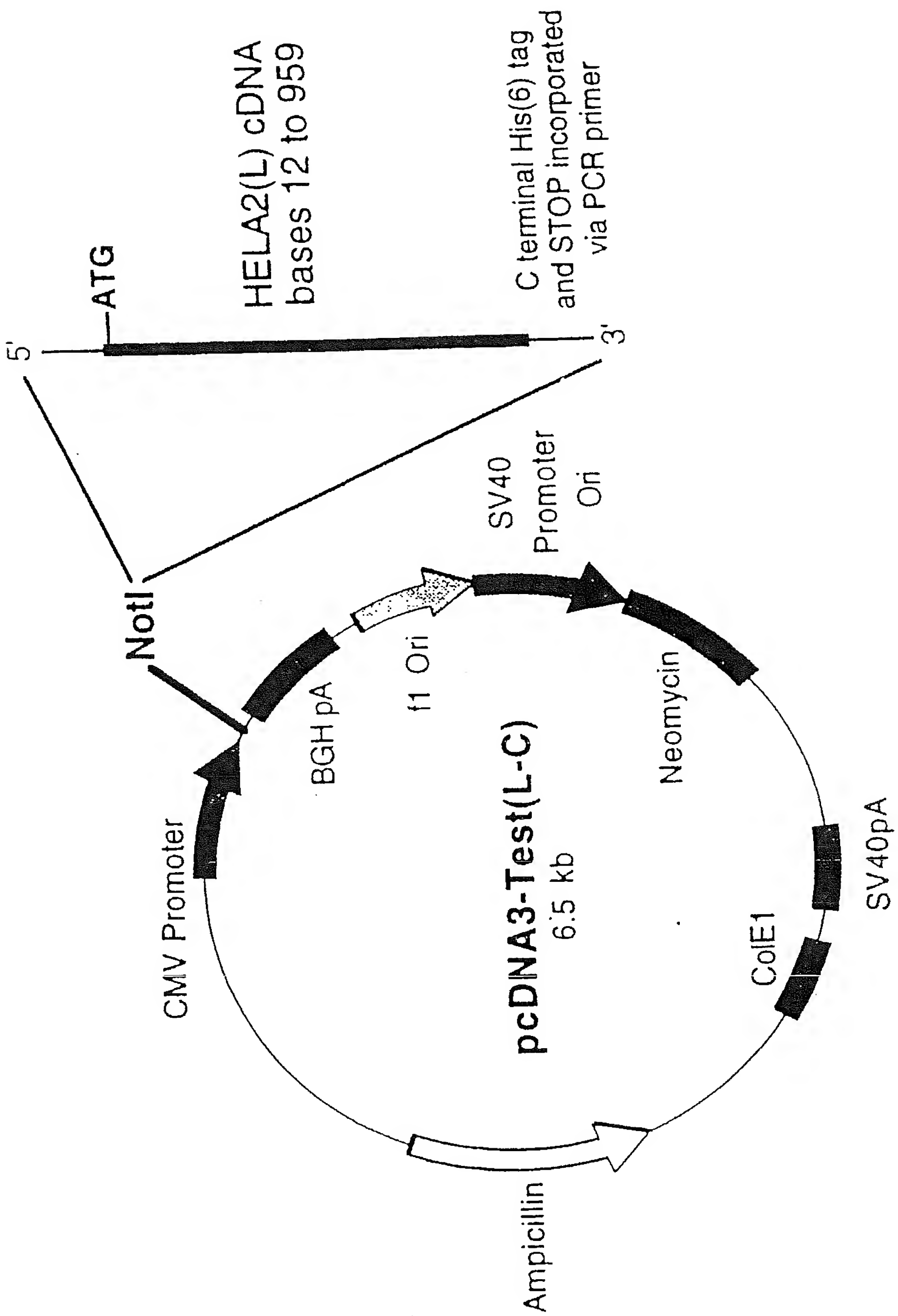

5'
ATG
HELA2(L) cDNA
bases 12 to 959
C terminal His(6) tag
and STOP incorporated
via PCR primer
3'
NotI
SV40
Promoter
Ori
f1 Ori
BGH pA
CMV Promoter
Neomycin
pcDNA3-Test(L-C)
6.5 kb
SV40pA
ColE1
Ampicillin

FIG 8(ii)

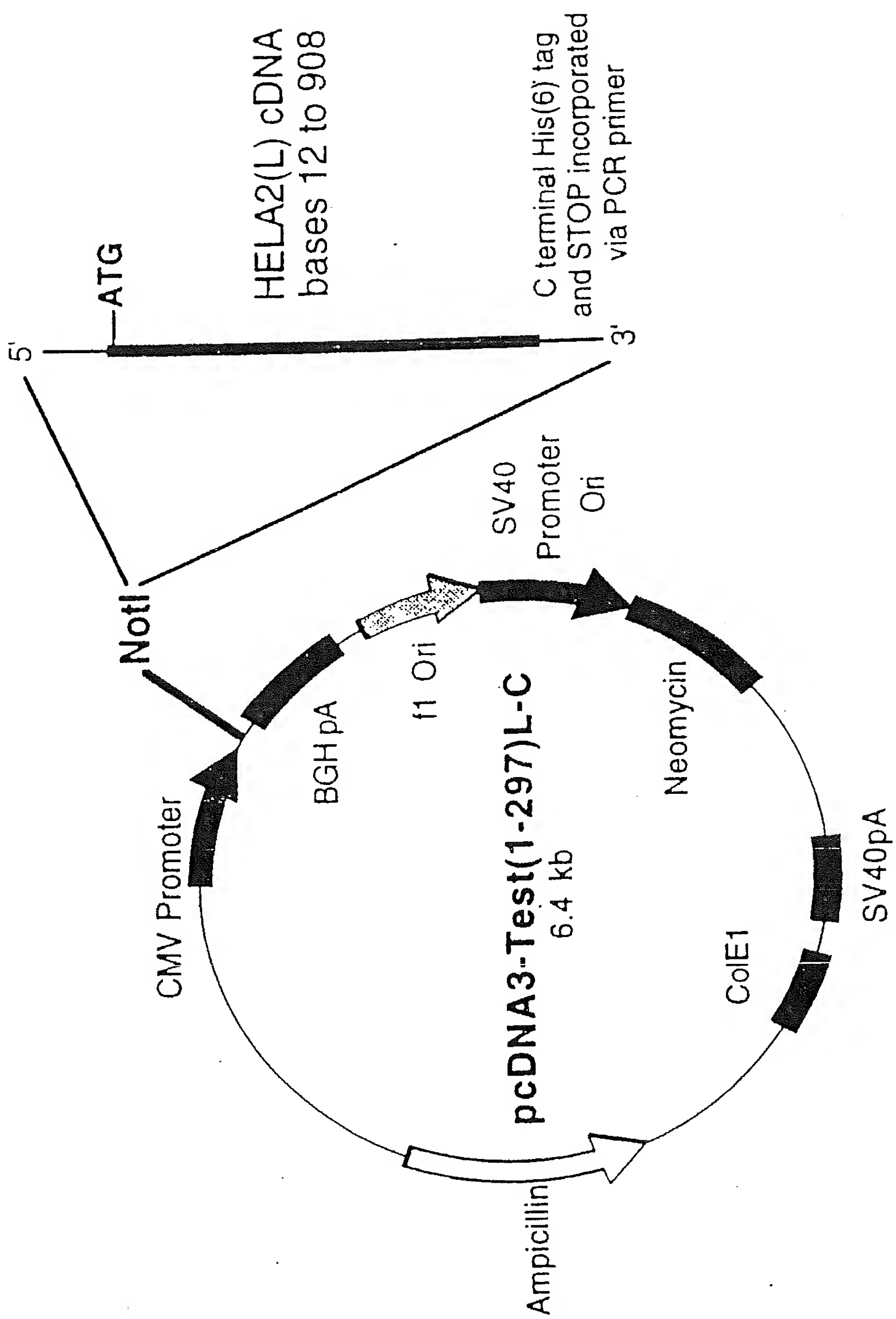

5'
ATG
HELA2(L) cDNA
bases 12 to 908
C terminal His(6) tag
and STOP incorporated
via PCR primer
3'
NotI
CMV Promoter
BGH pA
f1 Ori
SV40
Promoter
Ori
Neomycin
SV40pA
ColE1
Ampicillin
pcDNA3-Test(1-297)L-C
6.4 kb

FIG 8(iii)

FIG 9

FIG 9(i)	FIG 9(ii)
FIG 9(iii)	FIG 9(iv)

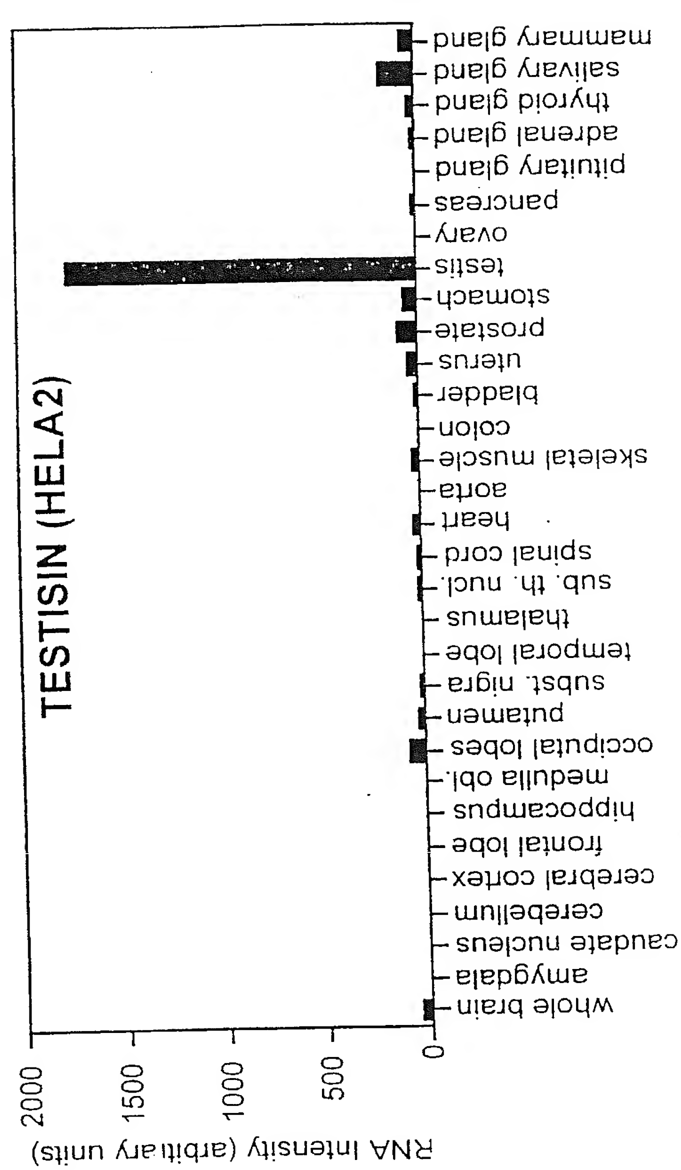

TESTISIN (HELA2)
RNA Intensity (arbitrary units)
2000
1500
1000
500
0
whole brain
amygdala
caudate nucleus
cerebellum
cerebral cortex
frontal lobe
hippocampus
medulla obl.
occiputal lobes
putamen
subst. nigra
temporal lobe
thalamus
sub. th. nucl.
spinal cord
heart
aorta
skeletal muscle
colon
bladder
uterus
prostate
stomach
testis
ovary
pancreas
pituitary gland
adrenal gland
thyroid gland
salivary gland
mammary gland

FIG 9(i)

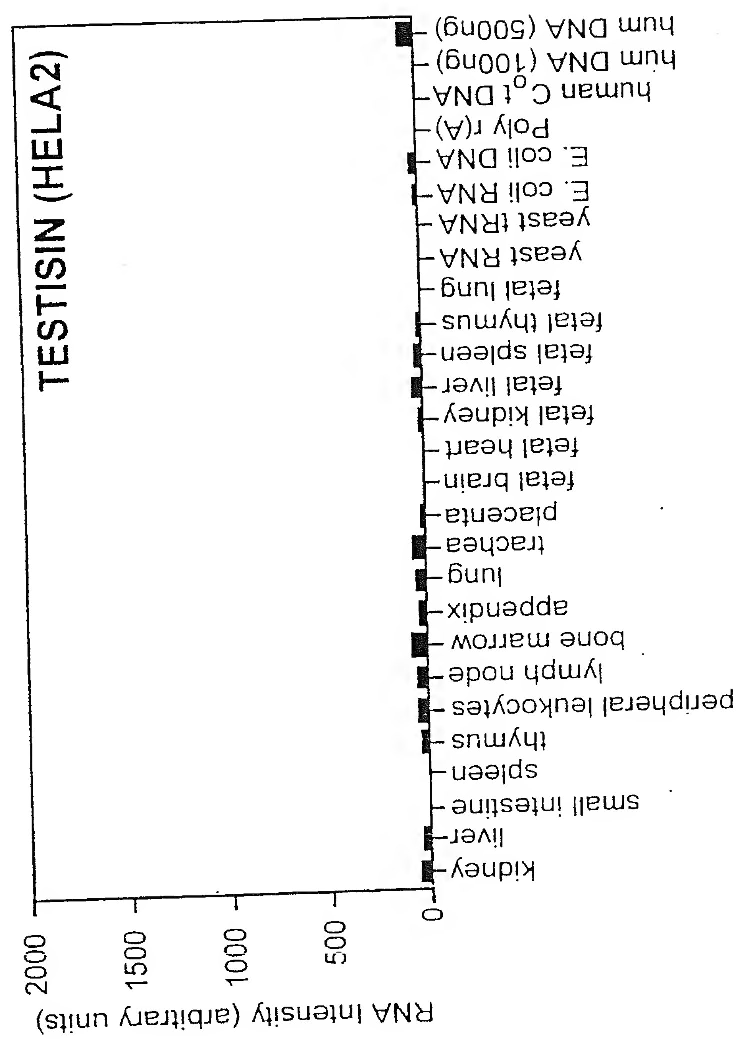

TESTISIN (HELA2)
RNA Intensity (arbitrary units)
2000
1500
1000
500
0
kidney
liver
small intestine
spleen
thymus
peripheral leukocytes
lymph node
bone marrow
appendix
lung
trachea
placenta
fetal brain
fetal heart
fetal kidney
fetal liver
fetal spleen
fetal thymus
fetal lung
yeast RNA
yeast tRNA
E. coli RNA
E. coli DNA
Poly r(A)
human C_ot DNA
hum DNA (100ng)
hum DNA (500ng)

FIG 9(ii)

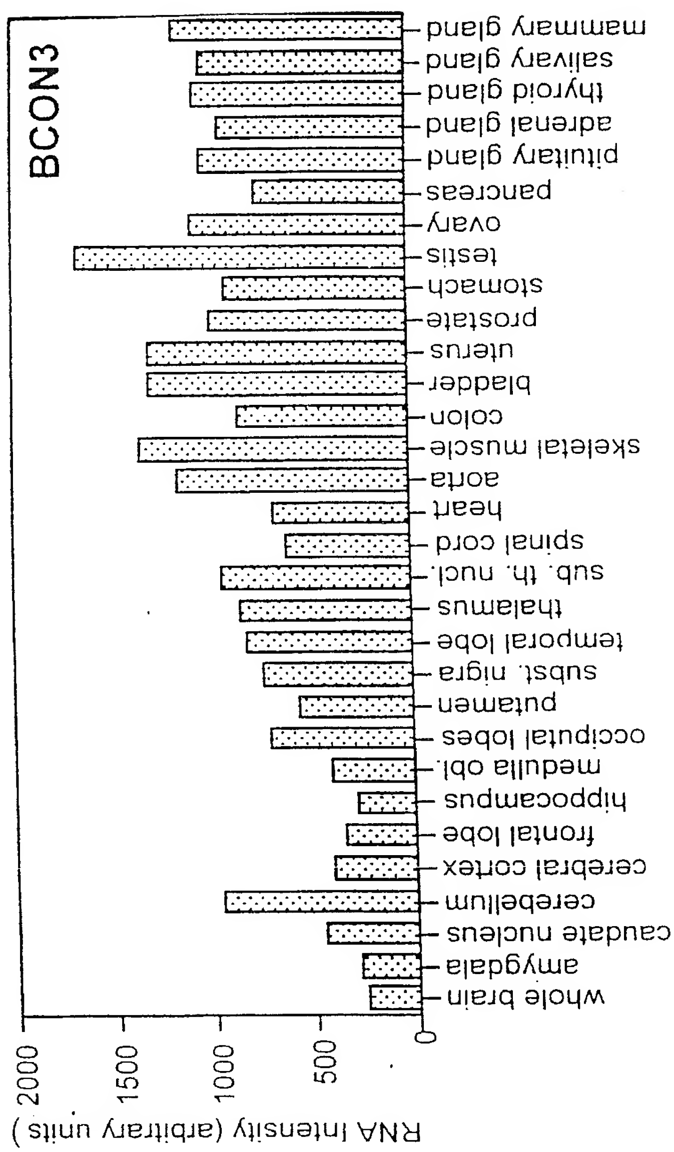

BCON3
RNA intensity (arbitrary units)
2000
1500
1000
500
0
whole brain
amygdala
caudate nucleus
cerebellum
cerebral cortex
frontal lobe
hippocampus
medulla obl.
occipital lobes
putamen
subst. nigra
temporal lobe
thalamus
sub. th. nucl.
spinal cord
heart
aorta
skeletal muscle
colon
bladder
uterus
prostate
stomach
testis
ovary
pancreas
pituitary gland
adrenal gland
thyroid gland
salivary gland
mammary gland

FIG 9(iii)

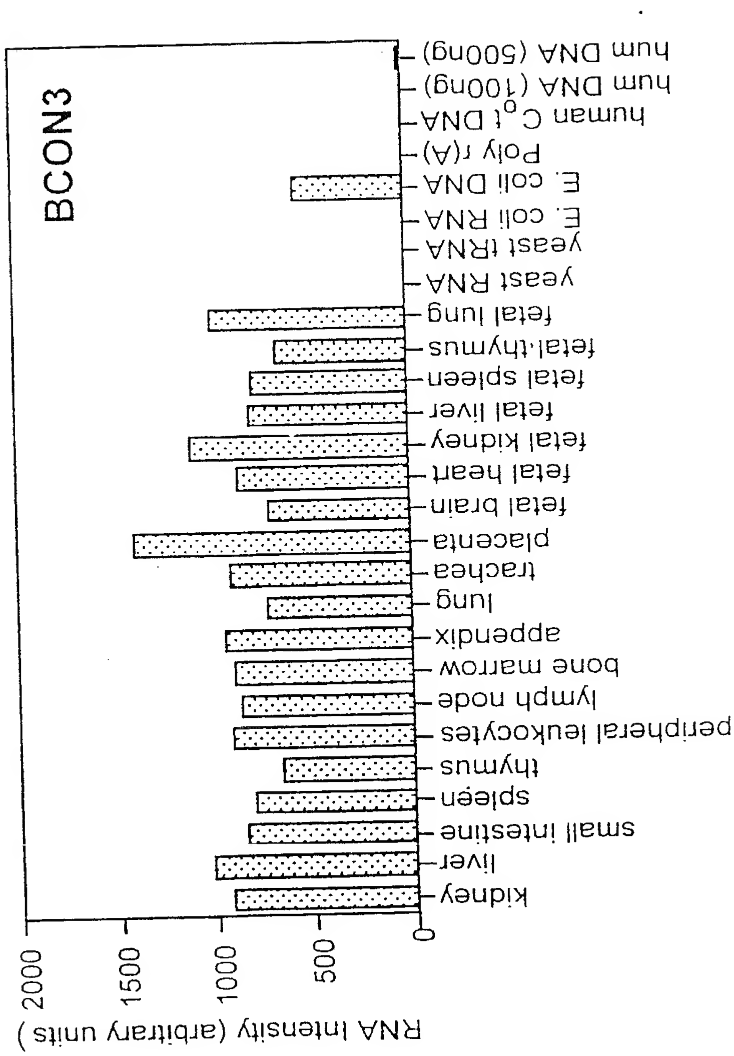

BCON3
RNA Intensity (arbitrary units)
2000
1500
1000
500
0
kidney
liver
small intestine
spleen
thymus
peripheral leukocytes
lymph node
bone marrow
appendix
lung
trachea
placenta
fetal brain
fetal heart
fetal kidney
fetal liver
fetal spleen
fetal.thymus
fetal lung
yeast RNA
yeast tRNA
E. coli RNA
E. coli DNA
Poly r(A)
human C_ot DNA
hum DNA (100ng)
hum DNA (500ng)

FIG 9(iv)

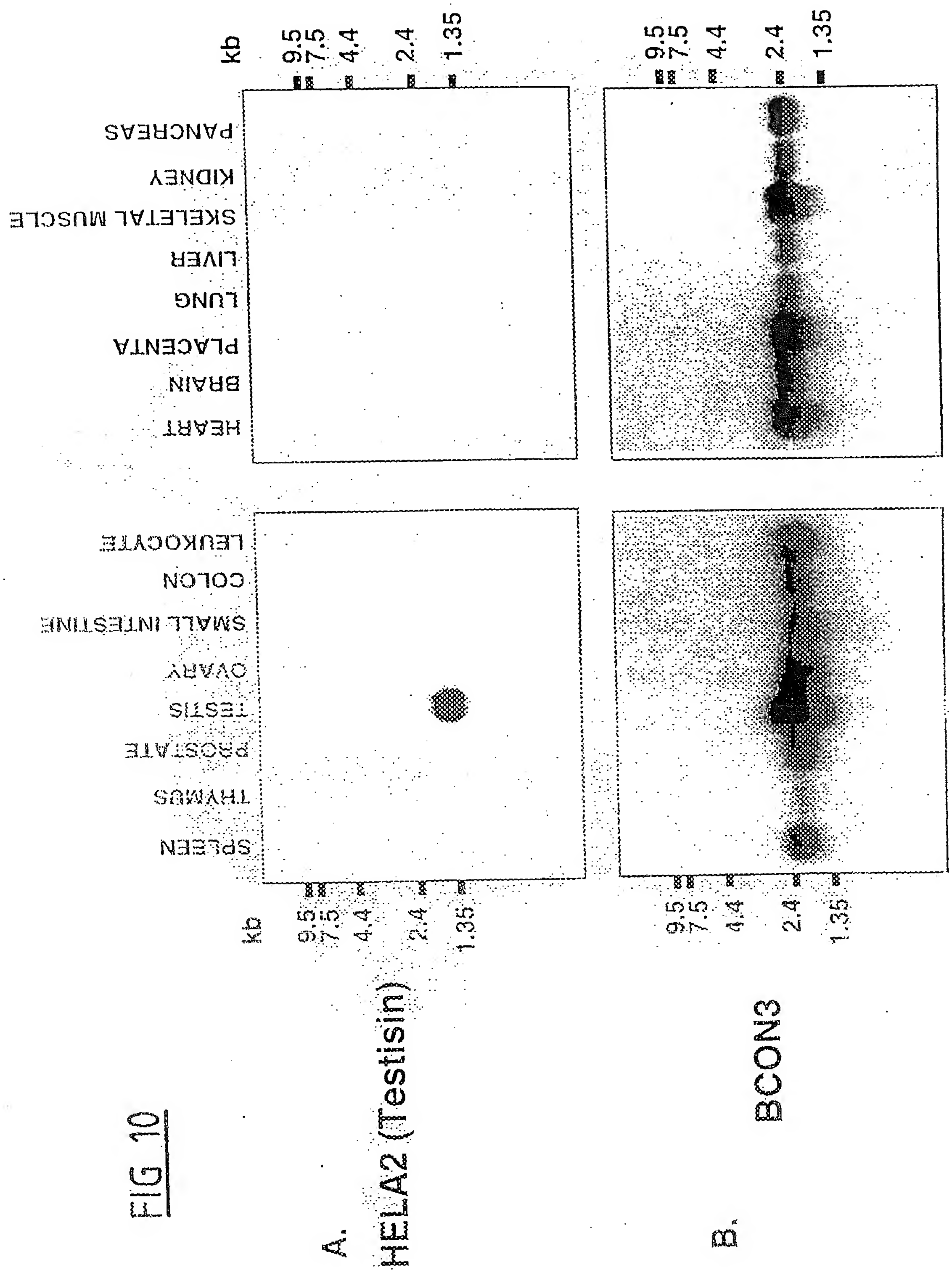

A.
HELA2 (Testisin)
B.
BCON3
kb
9.5
7.5
4.4
2.4
1.35
PANCREAS
KIDNEY
SKELETAL MUSCLE
LIVER
LUNG
PLACENTA
BRAIN
HEART
LEUKOCYTE
COLON
SMALL INTESTINE
OVARY
TESTIS
PROSTATE
THYMUS
SPLEEN

FIG 10

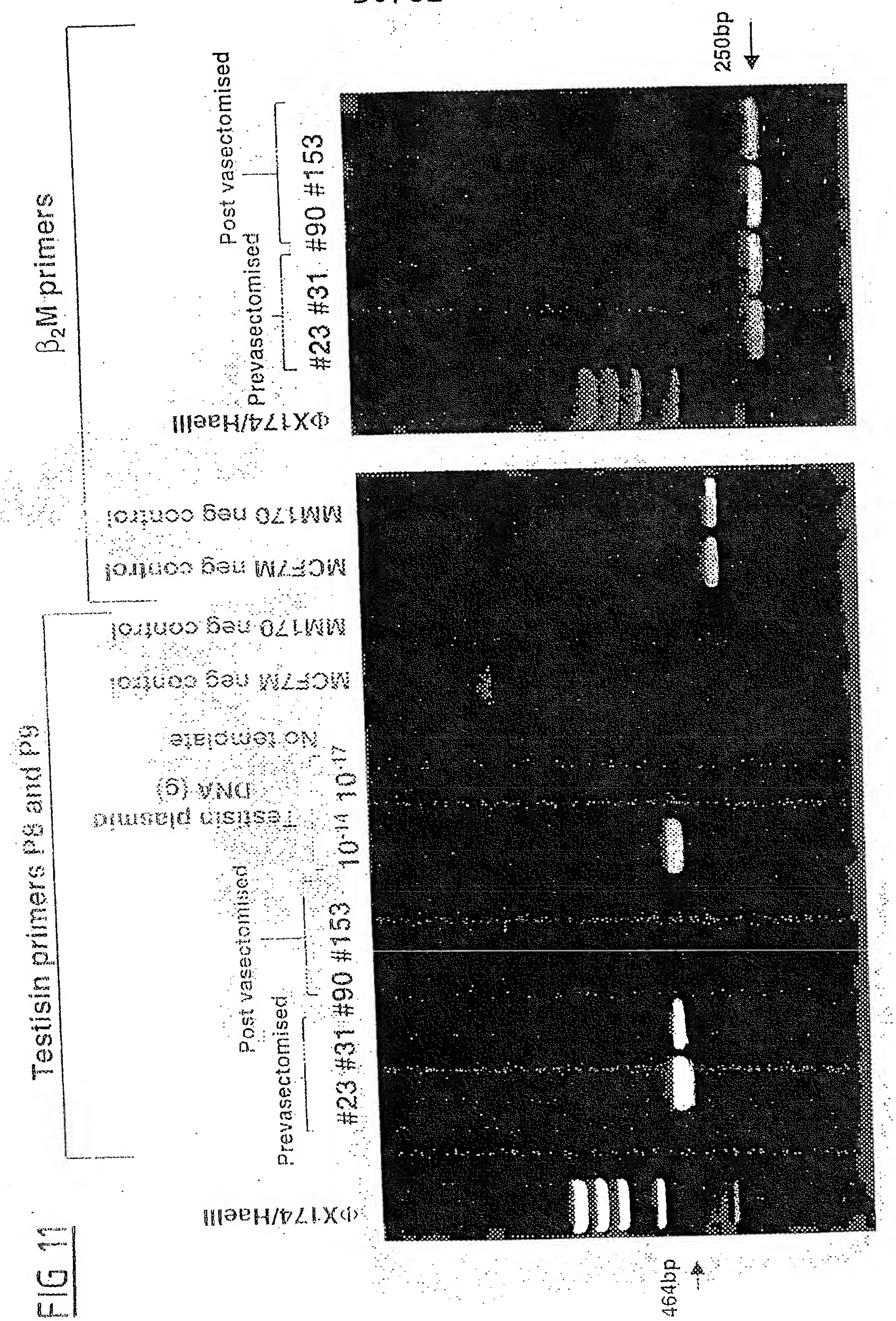

β2M primers
Testisin primers P8 and P9
Post vasectomised
Prevasectomised
#23 #31 #90 #153
ΦX174/HaeIII
MM170 neg control
MCF7M neg control
No template
Testisin plasmid DNA (g)
10-14 10-17
250bp
464bp

FIG 11

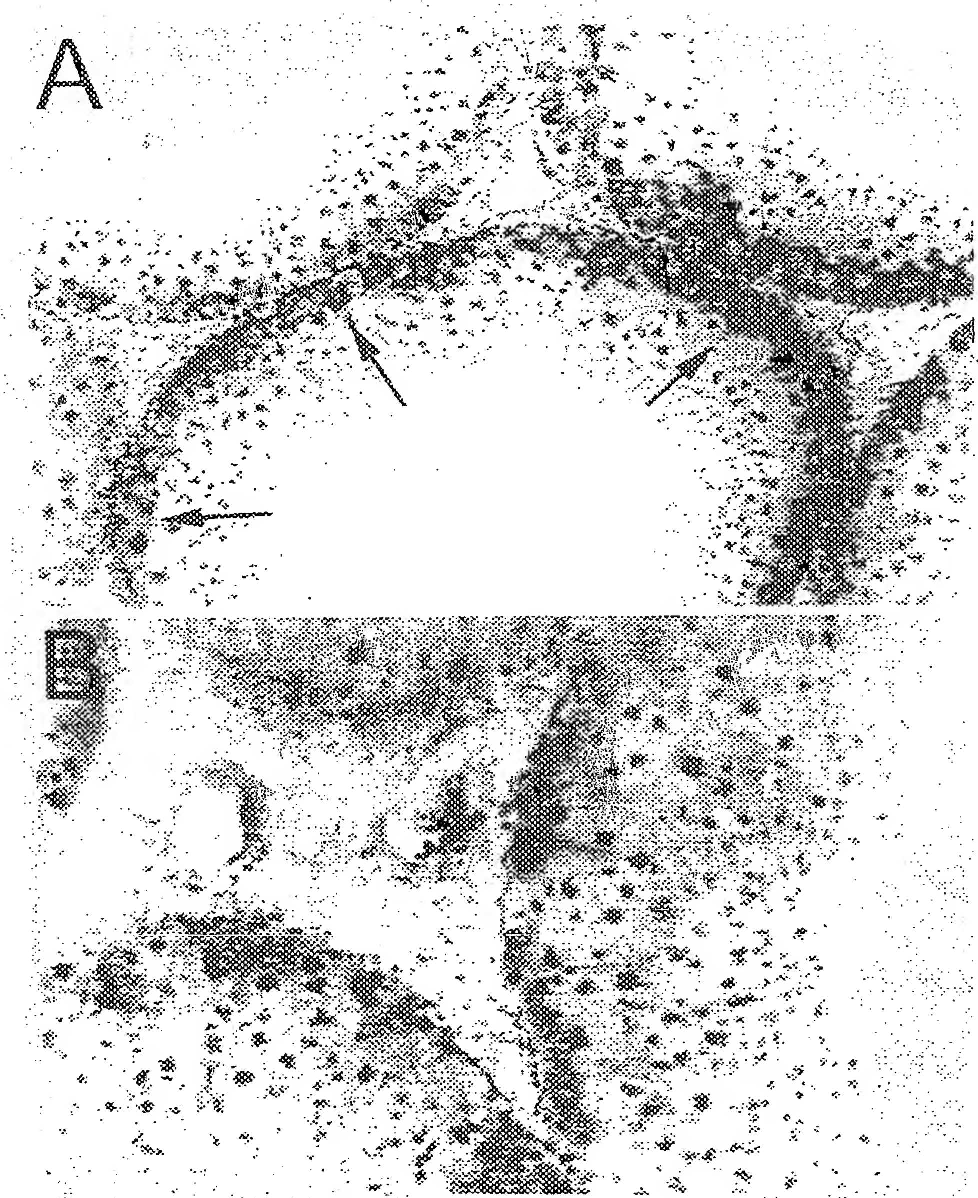


FIG 12

Testisin (HELA2) is located on human chromosome 16p13.3

A

FIG 13A

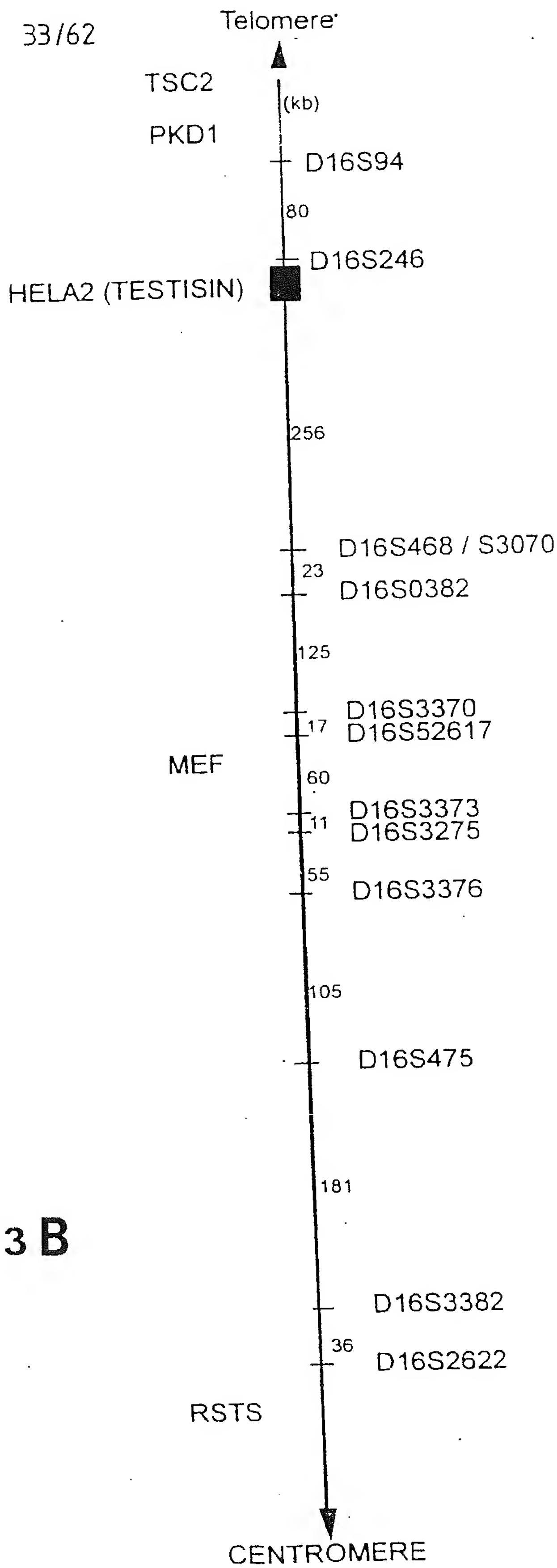
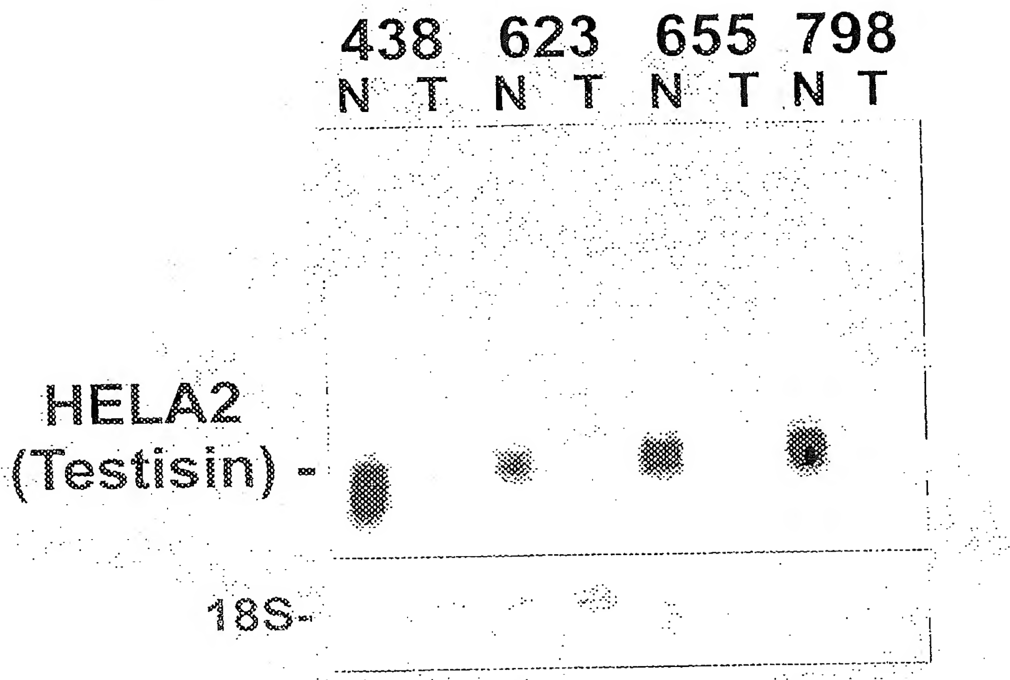

Telomere
TSC2
PKD1
(kb)
D16S94
80
D16S246
HELA2 (TESTISIN)
256
D16S468 / S3070
23
D16S0382
125
D16S3370
17
D16S52617
MEF
60
D16S3373
11
D16S3275
55
D16S3376
105
D16S475
181
D16S3382
36
D16S2622
RSTS
CENTROMERE

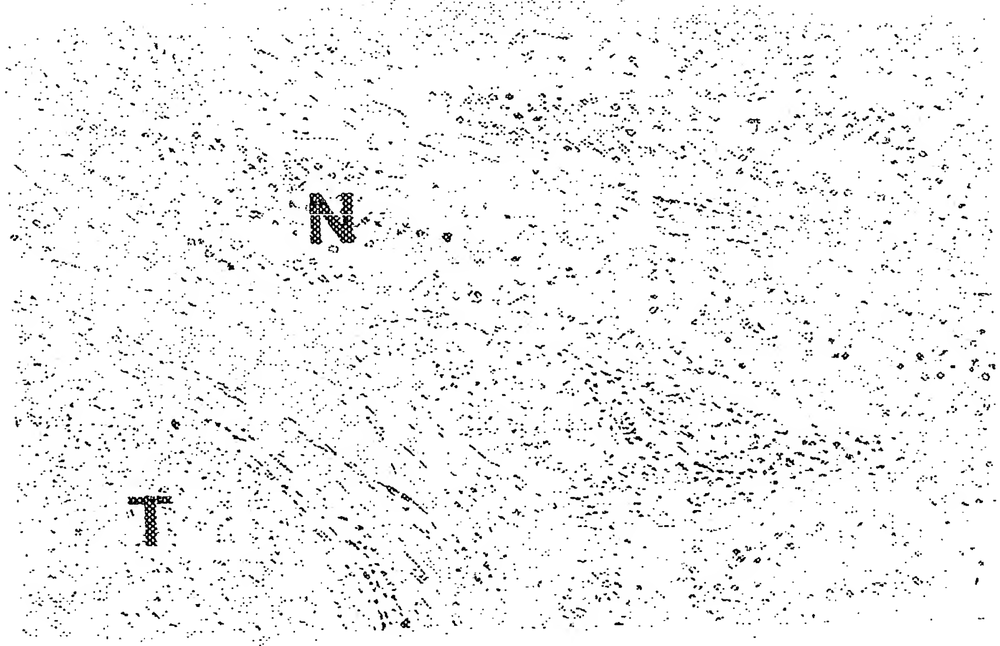
FIGURE 13 B

FIG 14

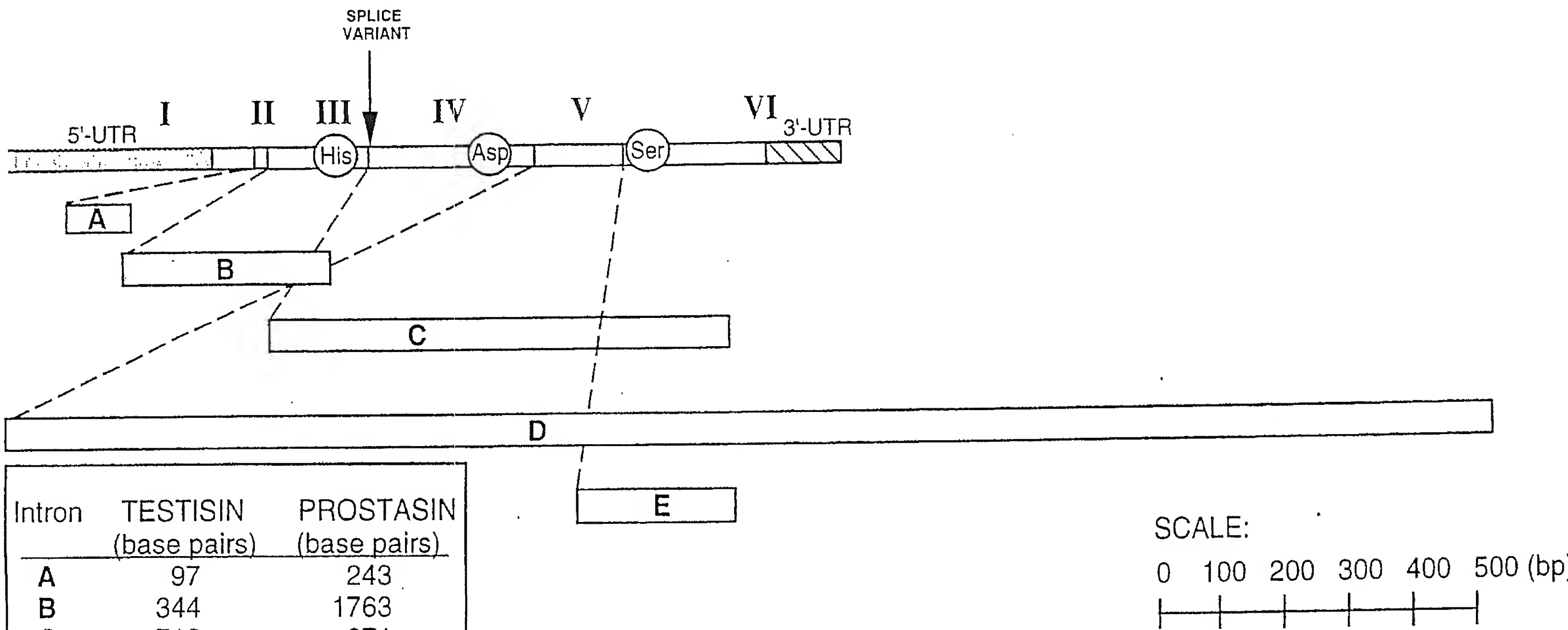
A. Northern Blot



B. Immunohistochemistry



TESTISIN INTRON/EXON BOUNDARIES AND SIZES

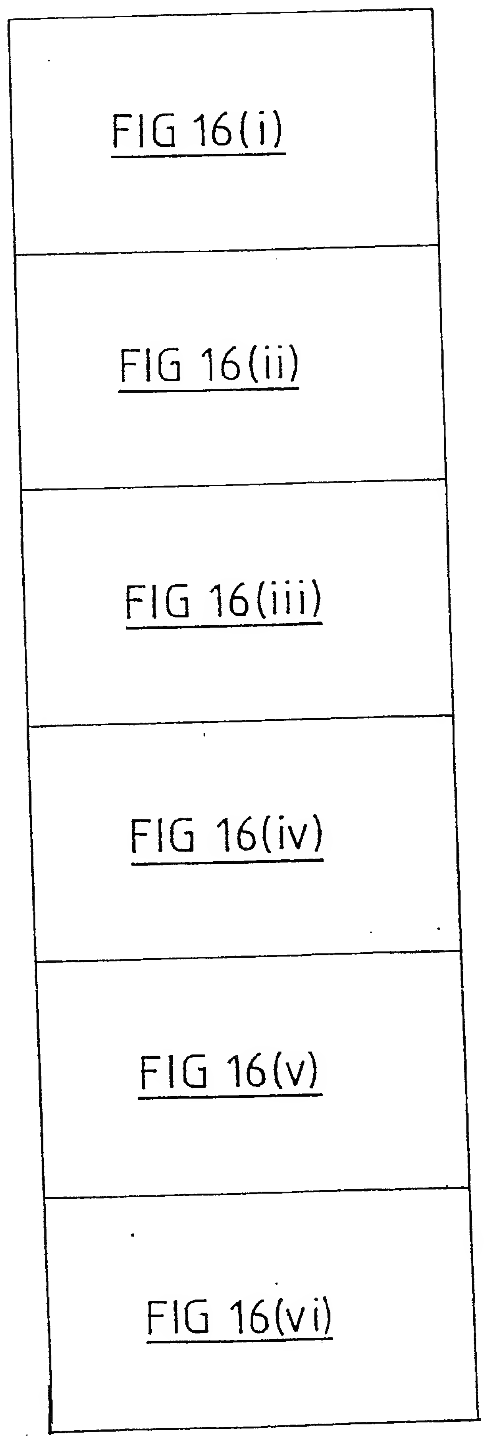


Intron	TESTISIN (base pairs)	PROSTASIN (base pairs)
A	97	243
B	344	1763
C	716	271
D	~2200	85
E	256	92

Exon	TESTISIN (base pairs)	PROSTASIN (base pairs)
I	>76	417
II	18	18
III	163	163
IV	284	272
V	168	167
VI	348	899

FIGURE 15

FIG 16



```
agtgagtctc ctgcctcagc ctcccaagta gctgggactt caggtgtgtg    50
ccaccatcct cagctaattt tttttttttt tttttttttg agaaggagtc   100
ttgctctgtc gcccaggctg gagtgcagtg gcgcgatctt ccaggcccca   150
ccgggccctc aggaaggcct tgcctacctg ctttaagggg actcctggct   200
cagggccagg cccctggtgc tggaggaggt ggtgggtgga gggcaggggg   250
caccaagcgg gcagccagga cccccgggct gcagacaaga aaaggactgt   300
                      /+1...EXON 1...
ggggtccacc gggtctgggc cACATCAAGG AATGTGGTTG AAGACCCGCC   350
CTTAGGAGCT GAAAGCCAGG GCGCTACCAG GCCTGAGAGG CCCCAAACAG   400
CCCTTGGGCC TGGTTTGGGA GGATTAAGCT GGAGCTCCCA ACCCGCCCTG   450
CCCCCAGGGG GCGACCCCGG GCCCGGCGCG AGAGGAGGCA GAGGGGGCGT   500
CAGGCCGCGG GAGAGGAGGC CATGGGCGCG CGCGGGGCGC TGCTGCTGGC   550
                                        /INTRON A...
GCTGCTGCTG GCTCGGGCTG GACTCAGGAA GCCGGgtgag ctcggggcgc   600
tgctggcggg atggggaggc gggggagcgg tggggaggac gggaggtgga   650
```

FIG 16(i)

```
                                   /EXON 2...
ggccgcgggg agtcacttct tgtctcccgc agAGTCGCAG GAGGCGGCGC      700
          /INTRON B...
CGTTATCAGg tagggcgccc aggacgcgcg attcctgcca gggccgttgg      750
gccgaggtgg acggggggcg gtgagggggt agagggggc ctttactgct       800
ctctcgcccc cgcccccggg atcgagaact ctgttggcgt ggaaagtaac      850
taacggacgc tggaggggga tgggcgggcc ctgcagagca cgtgggagga      900
tctccagtgt cacctacttc ctgctgcaca cacgcgaggg gaccctgggt      950
gggcaaaaac gtgctttccc ggacggggtt gaaggggaga aagggagagg     1000
tcgggcttgg ggggctgcct cccgcggctc agcagttcct ctgaccatcc     1050
     /EXON 3...
gagGACCATG CGGCCGACGG GTCATCACGT CGCGCATCGT GGGTGGAGAG     1100
GACGCCGAAC TCGGGCGTTG GCCGTGGCAG GGGAGCCTGC GCCTGTGGGA     1150
TTCCCACGTA TGCGGAGTGA GCCTGCTCAG CCACCGCTGG GCACTCACGG     1200
```

FIG 16(ii)

```
                    /INTRON  C...
CGGCGCACTG CTTTGAAACg tgagtggggg tgcgaacgga ggggtgcggg     1250
gacgggcagg aacagggctg gagggagtgc caccgaactt tacctctggt     1300
ctgatgccag acttgggcgt gaaagttgtg cgtggatgcg gcctggtgtt     1350
ctcctgagcc ccaggctgtg ctgcagccgg ttacacccac tccagttccc     1400
tttgggtctc ctggagggaa ccctgttcag gttattccag aatgttcttc     1450
cagaacattt ccacacactt ttgggtattc tctccctttt tctttcaacc     1500
caaagttcac cactgaccat cccaccctca tccccctcc tggtggacgg      1550
tgcggtacag tgtggggcac tgagccaagg ccagcacccc cgggccgctg     1600
tgtggactcc atcctgccaa tcccacattg gcgtggtgca tctccccatt     1650
cctccttggg ctgcatgggg gtgcccctgg aggccttggc tcaatgcaag     1700
gctccttggg acagctctgg gaggtgacaa gaccccaccc ttctgctgca     1750
ggagcaggtc ctaggacttt ggttgtggtc tgtctgggct ccttcatttc     1800
tgcaggggac cctgggtgtt agcaagtagc agcaacacca cagtttcccc     1850
tcctgcactg gaccccagtt gtgctcaggt agccagccct ccatccaggg     1900
```

FIG 16(iii)

```
                                          /EXON 4...
ccccctgactg ctctcttctc ttctgccagc tatagTGACC TTAGTGATCC    1950
CTCCGGGTGG ATGGTCCAGT TTGGCCAGCT GACTTCCATG CCATCCTTCT    2000
GGAGCCTGCA GGCCTACTAC ACCCGTTACT TCGTATCGAA TATCTATCTG    2050
AGCCCTCGCT ACCTGGGGAA TTCACCCTAT GACATTGCCT TGGTGAAGCT    2100
GTCTGCACCT GTCACCTACA CTAAACACAT CCAGCCCATC TGTCTCCAGG    2150
CCTCCACATT TGAGTTTGAG AACCGGACAG ACTGCTGGGT GACTGGCTGG    2200
                             /INTRON  D...
GGGTACATCA AAGAGGATGA GGgtgaggct ggggacaggc gggtcaggga    2250
ggaactgtct ttgttcacct gttcccctgc ataggcacaa tagcccccctg    2300
cttggtctgg gggtgcaggc tatgcccctc ttgcttgcag tctctcctca    2350
cctgccaggg cagggaccaa acacccagtt ctctcccttc caggggctgt    2400
gggggccaga aggagagtgt gagagggagg ccagtttggc gcaagcctgt    2450
gggtggtgcg gtggtggagg ggttctggag ggcttggcga cataaacctc    2500
atacttggat ttattcctgc atctttccac ctcccccagt gctcaccaat    2550
```

FIG 16(iv)

```
gccccaggca tca..............approx 1000 bp..........        3563
ccaggttgcc ccttccccca aggtctggct ttggatgctt atgtgaacac     ≈3613
cgttttaagt tgccttggcc ccttcctcgg ttcctttttg gctgaggaat     ≈3663
ctctccatgg ctgcaggcag ggccattgtt gccattctac agatagggaa     ≈3713
agtgcggctg ggggagctct gacagctgtc cctccccggg gccttctgtg     ≈3763
atgctgctga gggcctctgt tgtgctgggg tctgggttgg agctgggggt     ≈3813
aatggagatg aacctgccag gcacagtggg tgccccaggg cccccacccc     ≈3863
cgcagcctat gccatccctc catagagggg cctcaggttg ctgtctctct     ≈3913
                        /EXON 5...
ccttcccact atcgtccgca cagCACTGCC ATCTCCCCAC ACCCTCCAGG     ≈3963
AAGTTCAGGT CGCCATCATA AACAACTCTA TGTGCAACCA CCTCTTCCTC     ≈4013
AAGTACAGTT TCCGCAAGGA CATCTTTGGA GACATGGTTT GTGCTGGCAA     ≈4063
                              /INTRON  E...
TGCCCAAGGC GGGAAGGATG CCTGCTTCgt gagtgtcctt gccaccactc     ≈4113
ccagcccagg aaagcatcct gtgtccctgt gccttatttg accctcatgc     ≈4163
caaccccggg aggtggagac tgttgcccca ctctgcagat gcagaaacgg     ≈4213
```

FIG 16(v)

```
aggcttggct gctgccaggg ggaggaggag gatgtgcacc cagtctaccc     ≈4263
agccccatag cccttcccac tctcagcccc tcccctgccc cactcactct     ≈4313
                                 /EXON 6...
gccccaggct gacctcagcc ccgctgctcc ccagGGTGAC TCAGGTGGAC     ≈4363
CCTTGGCCTG TAACAAGAAT GGACTGTGGT ATCAGATTGG AGTCGTGAGC     ≈4413
TGGGGAGTGG GCTGTGGTCG GCCCAATCGG CCCGGTGTCT ACACCAATAT     ≈4463
CAGCCACCAC TTTGAGTGGA TCCAGAAGCT GATGGCCCAG AGTGGCATGT     ≈4513
CCCAGCCAGA CCCCTCCTGG CCGCTACTCT TTTTCCCTCT TCTCTGGGCT     ≈4563
CTCCCACTCC TGGGGCCGGT CTGAGCCTAC CTGAGCCCAT GCAGCCTGGG     ≈4613
GCCACTGCCA AGTCAGGCCC TGGTTCTCTT CTGTCTTGTT TGGTAATAAA     ≈4663
CACATTCCAG TTGATGCCTT GCAGGGCATT CTTCAaaagc agtggcttca     ≈4713
tggacagctc attctctctt gtgcagacag cctgtctgtg cccctggctc     ≈4763
acacccacat ctgttctgca ccatagaacc atctggttat ttcgatcaga     ≈4813
aagagaattg tgtgttgccc aggctggtct tgaacgccta gggtgtctcg     ≈4863
atc                                                        ≈4866
```

FIG 16(vi)

EXON III CACTGCTTTGAAAC**gt**gagtgggggtgcgaacggag
gggtgcggggacgggcaggaacagggctggagggagtgccaccga
actttacctctggtctgatgccagacttgggcgtgaaagttgtgc
gtggatgcggcctggtgttctcctgagccccaggctgtgctgcag
ccggttacacccactccagttccctttgggtctcctggagggaac
cctgttcaggttattccagaatgttcttccagaacatttccacac
acttttgggtattctctccctttttctttcaacccaaagttcacc
actgaccatcccaccctcatccccctcctggtggacggtgcggt
acagtgtggggcactgagccaaggccagcacccccgggccgctgt

...............INTRON C (716 BP)...............

gtggactccatcctgccaatcccacattggcgtggtgcatctccc
cattcctccttgggctgcatggggtgcccctggaggccttggct
caatgcaaggctccttgggacagctctgggaggtgacaagacccc
acccttctgctgcaggagcaggtcctagactttggttgtggtctg
tctgggctccttcatttctgcaggggaccctgggtgttagcaagt
agcagcaacaccacagtttcccctcctgcactggaccccagttgt
gctcaggtagccagccctccatccagggcccctgactgctctctt
ctcttctgcc**ag**ctat**ag**TGACCTTAGTGATCCC EXON IV

ALTERNATIVE
SPLICING EVENTS

SHORT ISOFORM

CACTGCTTTGAAACTGACCTTAGTGATCCC

H C F E T D L S D P

LONG ISOFORM

CACTGCTTTGAAAC**CTATAG**TGACCTTAGTGATCCC

H C F E T **Y** **S** D L S D P

FIGURE 17

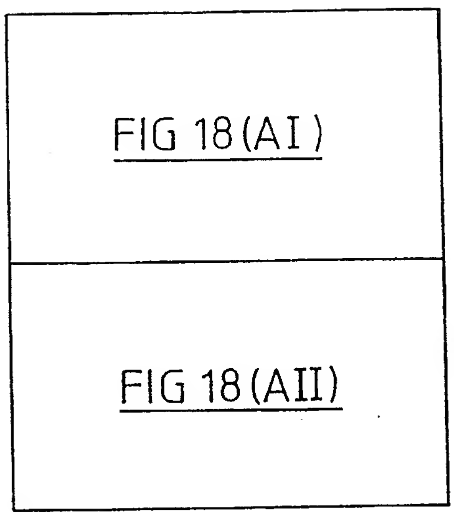


FIG 18(A)

FIGURE 18 (AI)

```
1   CGACCTATTGTCAGGGCCCTGCGGTCACAGGACCATCCCTTCCCGTATAGTGGGTGGCGA
     D  L  L  S  G  P  C  G  H  R  T  I  P  S  R  I  V  G  G  D  20

61  TGATGCTGAGCTTGGCCGCTGGCCGTGGCAAGGGAGCCTGCGTGTATGGGGCAACCACTT
     D  A  E  L  G  R  W  P  W  Q  G  S  L  R  V  W  G  N  H  L  40

121 ATGTGGCGCAACCTTGCTCAACCGCCGCTGGGTGCTTACAGCTGCCCACTGCTTCCAAAA
     C  G  A  T  L  L  N  R  R  W  V  L  T  A  A  H  C  F  Q  K  60

181 GGATAACGATCCTTTTGACTGGACAGTCCAGTTTGGTGAGCTGACTTCCAGGCCATCTCT
     D  N  D  P  F  D  W  T  V  Q  F  G  E  L  T  S  R  P  S  L  80

241 CTGGAACCTACAGGCCTATTCCAACCGTTACCAAATAGAAGATATTTTCCTGAGCCCCAA
     W  N  L  Q  A  Y  S  N  R  Y  Q  I  E  D  I  F  L  S  P  K 100

301 GTACTCGGAGCAGTATCCCAATGACATAGCCCTGCTGAAGCTGTCATCTCCAGTCACCTA
     Y  S  E  Q  Y  P  N  D  I  A  L  L  K  L  S  S  P  V  T  Y 120

361 CAATAACTTCATCCAGCCCATCTGCCTCCTGAACTCCACGTACAAGTTTGAGAACCGAAC
     N  N  F  I  Q  P  I  C  L  L  N  S  T  Y  K  F  E  N  R  T 140

421 TGACTGCTGGGTGACCGGCTGGGGGCTATTGGAGAAGATGAGAGTCTGCCATCTCCCAA
     D  C  W  V  T  G  W  G  A  I  G  E  D  E  S  L  P  S  P  N 160
```

FIGURE 18 (AII)

```
481 CACTCTCCAGGAAGTGCAGGTAGCTATTATCAACAACAGCATGTGTAACCATATGTACAA
     T  L  Q  E  V  Q  V  A  I  I  N  N  S  M  C  N  H  M  Y  K 180

541 AAAGCCAGACTTCCGCACGAACATCTGGGGAGACATGGTTTGCGCTGGCACTCCTGAAGG
     K  P  D  F  R  T  N  I  W  G  D  M  V  C  A  G  T  P  E  G 200

601 TGGCAAGGATGCCTGCTTTGGTGACTCGGGAGGACCCTTGGCCTGCGACCAGGATACGGT
     G  K  D  A  C  F  G  D  S  G  G  P  L  A  C  D  Q  D  T  V 220

661 GTGGTATCAGGTTGGAGTTGTGAGCTGGGGAATAGGCTGTGGTCGCCCCAATCGCCCTGG
     W  Y  Q  V  G  V  V  S  W  G  I  G  C  G  R  P  N  R  P  G 240

721 AGTCTATACCAACATCAGTCATCACTACAACTGGATCCAGTCAACCATGATCCGCAATGG
     V  Y  T  N  I  S  H  H  Y  N  W  I  Q  S  T  M  I  R  N  G 260

781 GCTGCTCAGGCCTGACCCAGTCCCCTTGCTACTGTTTCTTACTCTGGCCTGGGCTTCCTC
     L  L  R  P  D  P  V  P  L  L  L  F  L  T  L  A  W  A  S  S 280
841 TTTGCTGAGGCCTGCCTGAGCCCACACGTGTACGTCACACCTGTGAGGTCAGGGTGTGTC
     L  L  R  P  A                                               285

901 TCTTTTGTATCTTGCTTGCTAATAAACCTGTTAATATTTAAAAAAAAAAAAAAAAAAAA
```

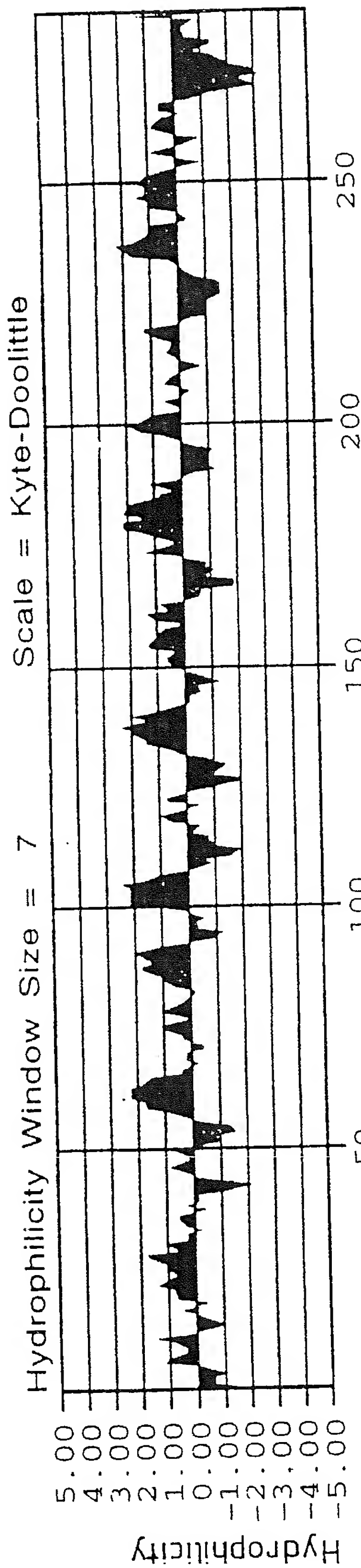

Hydrophilicity Window Size = 7 Scale = Kyte-Doolittle
5.00
4.00
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
Hydrophilicity
50
100
150
200
250

FIG 18B

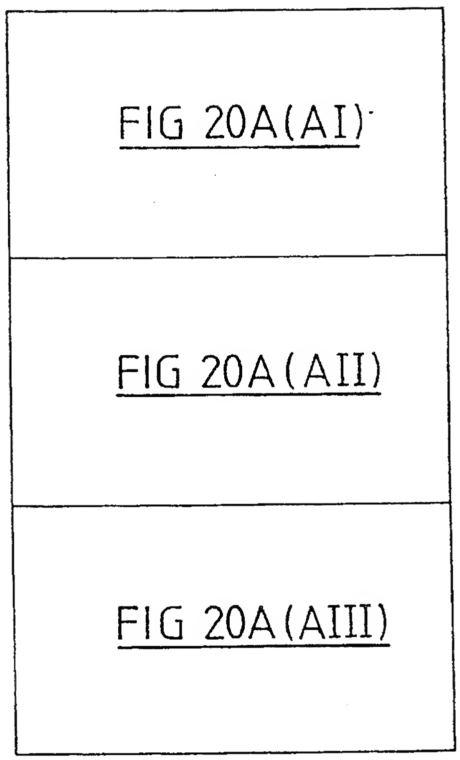


FIG 20A(A)

FIGURE 20A (AI)

```
     CTGAACCGGGTTGTGGGCGGCGAGGACAGCACTGACAGCGAGTGGCCCTGGATCGTGAGC  60
1     L  N  R  V  V  G  G  E  D  S  T  D  S  E  W  P  W  I  V  S

     ATCCAGAAGAATGGGACCCACCACTGCGCAGGTTCTCTGCTCACCAGCCGCTGGGTGATC 120
21    I  Q  K  N  G  T  H  H  C  A  G  S  L  L  T  S  R  W  V  I

     ACTGCTGCCCACTGTTTCAAGGACAACCTGAACAAACCATACCTGTTCTCTGTGCTGCTG 180
41    T  A  A  H  C  F  K  D  N  L  N  K  P  Y  L  F  S  V  L  L

     GGGGCCTGGCAGCTGGGGAACCCTGGCTCTCGGTCCCAGAAGGTGGGTGTTGCCTGGGTG 240
61    G  A  W  Q  L  G  N  P  G  S  R  S  Q  K  V  G  V  A  W  V

     GAGCCCCACCCTGTGTATTCCTGGAAGGAAGGTGCCTGTGCAGACATTGCCCTGGTGCGT 300
81    E  P  H  P  V  Y  S  W  K  E  G  A  C  A  D  I  A  L  V  R

     CTCGAGCGCTCCATACAGTTCTCAGAGCGGGTCCTGCCCATCTGCCTACCTGATGCCTCT 360
101   L  E  R  S  I  Q  F  S  E  R  V  L  P  I  C  L  P  D  A  S

     ATCCACCTCCCTCCAAACACCCACTGCTGGATCTCAGGCTGGGGGAGCATCCAAGATGGA 420
121   I  H  L  P  P  N  T  H  C  W  I  S  G  W  G  S  I  Q  D  G
```

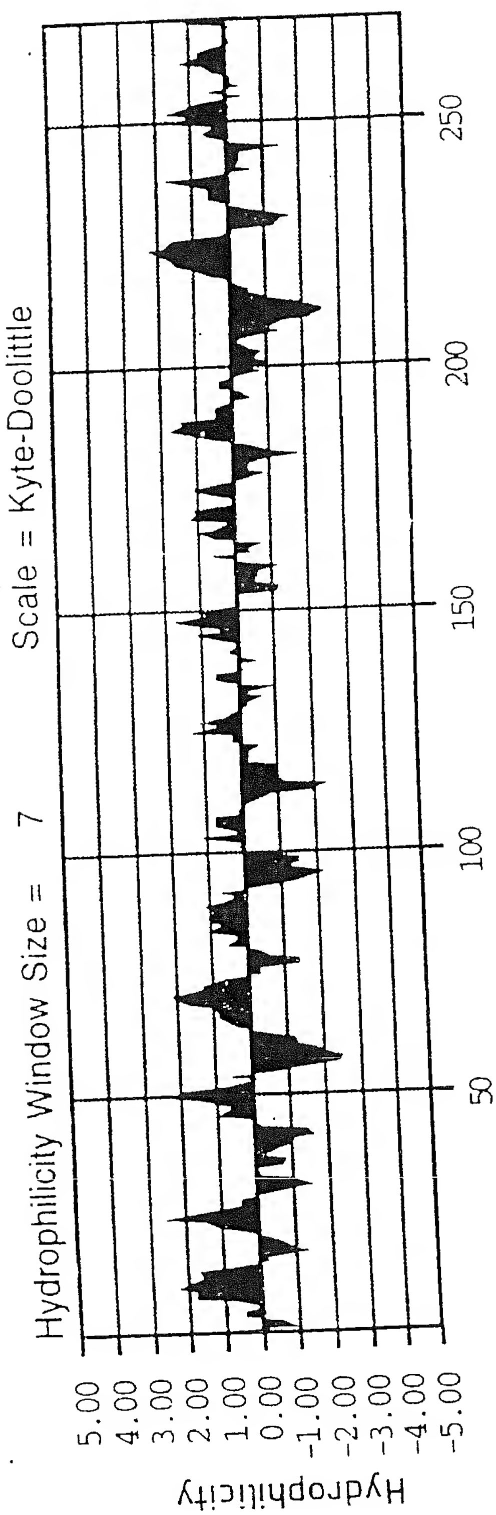
FIGURE 20A (AII)

```
    GTTCCCTTGCCCCACCCTCAGACCCTGCAGAAGCTGAAGGTTCCTATCATCGACTCGGGAA 480
141  V  P  L  P  H  P  Q  T  L  Q  K  L  K  V  P  I  I  D  S  E

    GTCTGCAGCCATCTGTACTGGCGGGGAGCAGGACAGGGACCCATCACTGAGGACATGCTG  540
161  V  C  S  H  L  Y  W  R  G  A  G  Q  G  P  I  T  E  D  M  L

    TGTGCCGGCTAACTTGGAGGGGGAGCGGGATGCTTGTCTGGGCGACTCCGGGGGCCCCCTC 600
181  C  A  G  Y  L  E  G  E  R  D  A  C  L  G  D  S  G  G  P  L

    ATGTGCCAGGTGGACGGCGCCTGGCTGCTGGCCGGCATCATCAGCTGGGGCGAGGGCTGT  660
201  M  C  Q  V  D  G  A  W  L  L  A  G  I  I  S  W  G  E  G  C

    GCCGAGCGCAACAGGCCCGGGGTCTACATCAGCCTCTCTGCGCACCGCTCCTGGGTGGAG  720
221  A  E  R  N  R  P  G  V  Y  I  S  L  S  A  H  R  S  W  V  E

    AAGATCGTGCAAGGGGTGCAGCTCCGCGGGCGCGCTCAGGGGGGTGGGGCCCTCAGGGCA  780
241  K  I  V  Q  G  V  Q  L  R  G  R  A  Q  G  G  G  A  L  R  A

    CCGAGCCAGGGCTCTGGGGCCGCCGCGCTCCTAGGGGCCCAGCGGGACGCGGGGCTCGG 840
261  P  S  Q  G  S  G  A  A  A  R  S

    ATCTGAAAGGCGGCCAGATCCACATCTGGATCTGGATCTGCGGCGGCCTCGGGCGGTTTC  900
    CCCCGCCGTAAATAGGCTCATCTACCTCTACCTCTGGGGGCCCGGACGGCTGCTGCGGAA  960
```

FIGURE 20A (AIII)

```
AGGAAACCCCCTCCCCGACCCGCCCGACGGCCTCAGGCCCCGCCTCCAAGGCATCAGGCC 1020
CCGCCCAACGGCCTCATGTCCCCGCCCCCACGACTTCCGGCCCCGCCCCGGGCCCCAGCG 1080
CTTTTGTGTATATAAATGTTAATGATTTTTATAGGTATTTGTAACCCTGCCACATATCT  1140
TATTTATTCCTCCAATTTCAATAAA
```

FIG 20A (B)


Hydrophilicity Window Size = 7 Scale = Kyte-Doolittle
Hydrophilicity
5.00
4.00
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
50
100
150
200
250

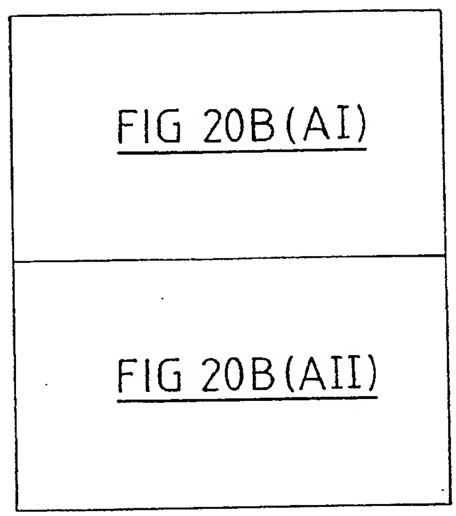


FIG 20B(A)

FIGURE 20B (AI)

AATGCGGCCACTCCAAGGAGGCCGGGAGGATTGTGGGAGGCCAACACACCCAGGAAGGAC 60
1 $\boxed{C}$ G H S K E A G R▽I V G G Q D T Q E G

GCTGGCCGTGGCAGGTTGGCCTGTGGTTGACCTCAGTGGGCATGTATGTGGGGCTCCC 120
21 R W P W Q V G L W L T S V G H V $\boxed{C}$ G G S

TCATCCACCCACGCTGGGTGCTCACAGCCGCCCACTGCTTCCTGAGGTCTGAGGATCCCG 180
41 L I H P R W V L T A A H $\boxed{C}$ F L R S E D P

GGCTCTACCATGTTAAAGTCGGAGGGCTGACACCCTCACTTTCAGAGCCCCACTCGGCCT 240
61 G L Y H V K V G G L T P S L S E P H S A

TGGTGGCTGTGAGGAGGCTCCTGGTCCACTCCTCATACCATGGACCACCACCAGCGGGG 300
81 L V A V R R L L V H S S Y H G T T T S G

ACATTGCCCTGATGGAGCTGGACTCCCCCTTGCAGGCCTCCCAGTTCAGCCCCATCTGCC 360
101 D I A L M E L D S P L Q A S Q F S P I $\boxed{C}$

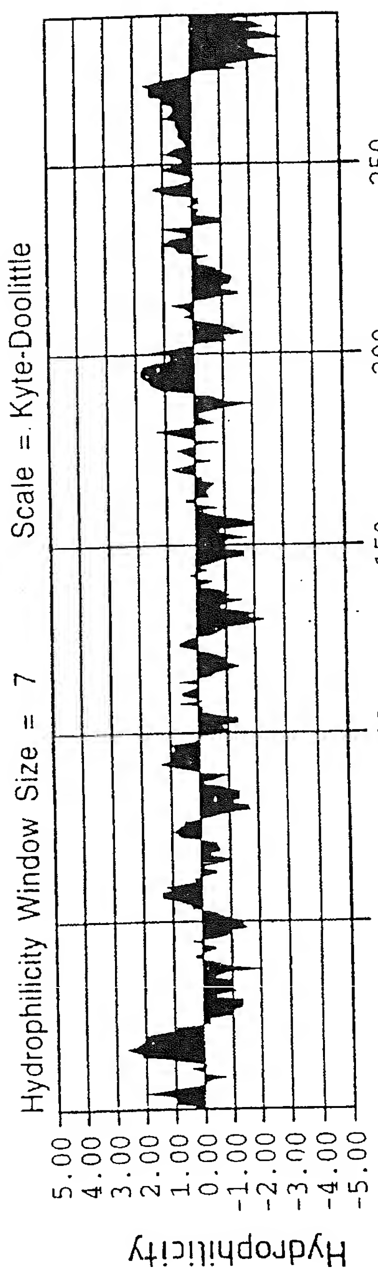
TCCCAGGACCCCAGACCCCCCTCGCCATTGGGACCGTGTGCTGGGTAAACGGCTGGGGG 420
121 L P G P Q T P L A I G T V $\boxed{C}$ W V N G L G

TCCACTCAGGAGAGGCCCTGGCGAGTGTCCTTCAGGAGGTGGCTGTGCCCCTCCTGGACT 480
141 V H S G E A L A S V L Q E V A V P L L D

FIGURE 20B (AII)

```
    CGAACATGTGTGAGCTGATGTACCACCTAGGAGAGCCCAGCCTGGCTGGCCAGCGCCTCA 540
161 S  N  M  C  E  L  M  Y  H  L  G  E  P  S  L  A  G  Q  R  L

    TCCAGGACGACATGCTCTGTGCTGGCTCTGTCCAGGGCAAGAAAGACTCCTGCCAGGGTG 600
181 I  Q  D  D  M  L  C  A  G  S  V  Q  G  K  K  D  S  C  Q  G

    ACTCCGGGGGGCCGCTGGTCTGCCCCATCAATGATACGTGGATCCAGGCCGGCATTGTGA 660
201 D  S  G  G  P  L  V  C  P  I  N  D  T  W  I  Q  A  G  I  V

    GCTGGGGATTCGGCTGTGCCCGGCCTTTCCGGCCTGGTGTCTACACCCAGGTGCTAAGCT 720
221 S  W  G  F  G  C  A  R  P  F  R  P  G  V  Y  T  Q  V  L  S

    ACACAGACTGGATTCAGAGAACCCTGGCTGAATCTCACTCAGGCATGTCTGGGGCCCGCC 780
241 Y  T  D  W  I  Q  R  T  L  A  E  S  H  S  G  M  S  G  A  R

    CAGGTGCCCCAGGATCCCACTCAGGCACCTCCAGATCCCACCCAGTGCTGCTGCTTGAGC 840
261 P  G  A  P  G  S  H  S  G  T  S  R  S  H  P  V  L  L  L  E

    TGTTGACCGTATGCTTGCTTGGGTCCCTGTGAACCATGAGCCATGGAGTCCGGGATCCCC 900
281 L  L  T  V  C  L  L  G  S  L

    TTTCTGGTAGGATTGATGGAATCTAATAATAAA
```

FIG 20B(B)


Hydrophilicity Window Size = 7 Scale = Kyte-Doolittle
Hydrophilicity
5.00
4.00
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
50
100
150
200
250

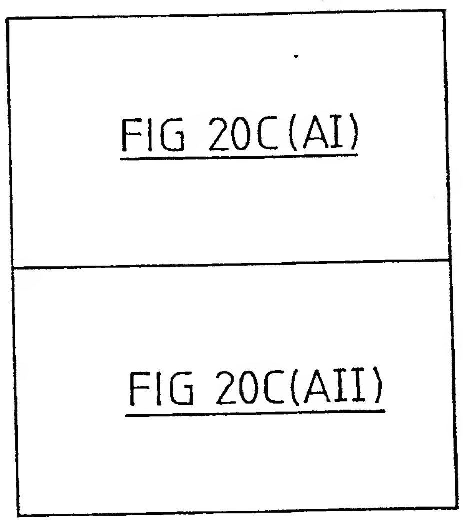


FIG 20C(A)

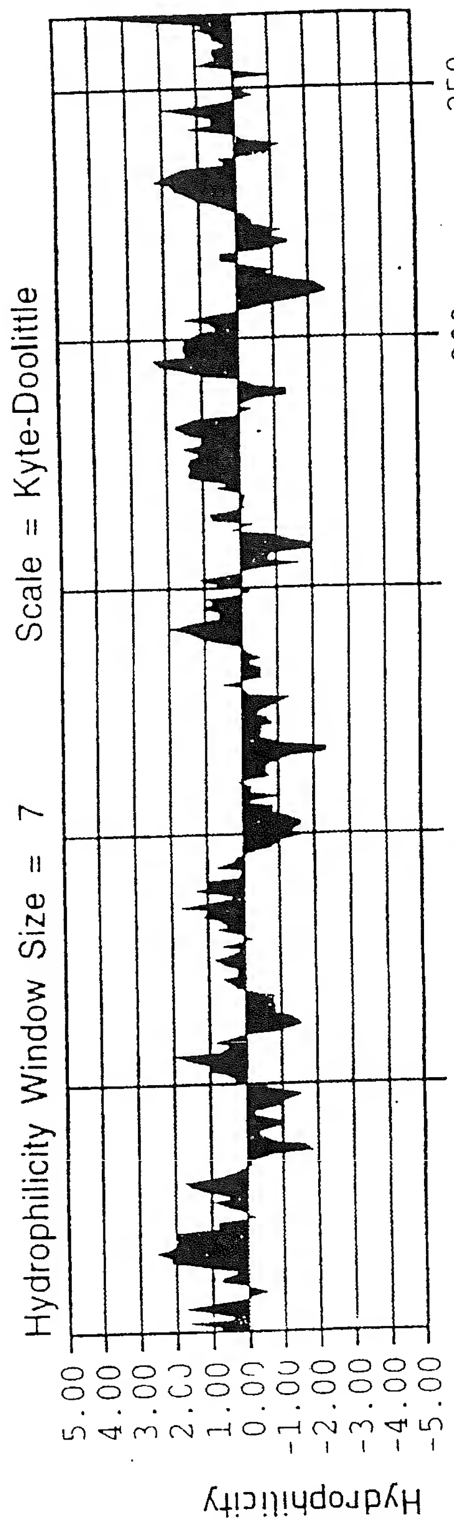
FIGURE 20C (AI)

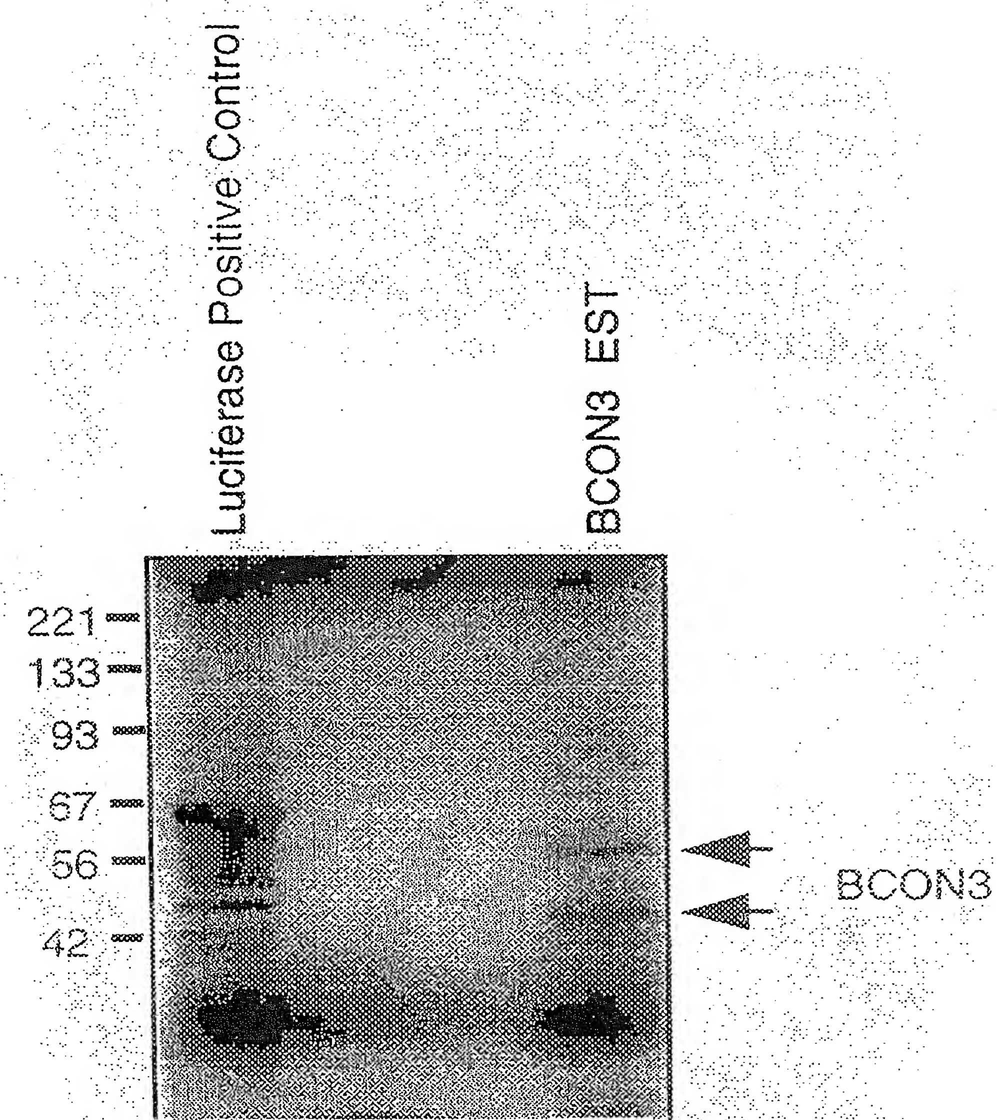
```
    CCTGTGGTCGCCCCAGGATGCTGAACCGAATGGTGGGCGGGCAGGACACGCAGGAGGGCG  60
1    C  G  R  P  R  M  L  N  R▽M  V  G  G  Q  D  T  Q  E  G

    AGTGGCCCTGGCAAGTCAGCATCCAGCGCAACGGAAGCCACTTCTGCGGGGGCAGCCTCA 120
21  E  W  P  W  Q  V  S  I  Q  R  N  G  S  H  F  C  G  G  S  L

    TCGCGGAGCAGTGGGTCCTGACGGCTGCGCACTGCTTCCGCAACACCTCTGAGACGTCCC 180
41  I  A  E  Q  W  V  L  T  A  A  H  C  F  R  N  T  S  E  T  S

    TGTACCAGGTCCTGCTGGGGGCAAGGCAGCTAGTGCAGCCGGGACCACACGCTATGTATG 240
61  L  Y  Q  V  L  L  G  A  R  Q  L  V  Q  P  G  P  H  A  M  Y

    CCCGGGTGAGGCAGGTGGAGAGCAACCCCCTGTACCAGGGCACGGCCTCCAGCGCTGACG 300
81  A  R  V  R  Q  V  E  S  N  P  L  Y  Q  G  T  A  S  S  A  D

    TGGCCCTGGTGGAGCTGGAGGCACCAGTGCCCTTCACCAATTACATCCTCCCCGTGTGCC 360
101 V  A  L  V  E  L  E  A  P  V  P  F  T  N  Y  I  L  P  V  C

    TGCCTGACCCCTCGGTGATCTTTGAGACGGGCATGAACTGCTGGGTCACTGGCTGGGGCA 420
121 L  P  D  P  S  V  I  F  E  T  G  M  N  C  W  V  T  G  W  G

    GCCCCAGTGAGGAAGACCTCCTGCCCGAACCGCGGATCCTGCAGAAACTCGCTGTGCCCA 480
141 S  P  S  E  E  D  L  L  P  E  P  R  I  L  Q  K  L  A  V  P
```

FIGURE 20C (AII)

```
    TCATCGACACACCCAAGTGCAACCTGCTCTACAGCAAAGACACCGAGTTTGGCTACCAAC 540
161 I  I  D  T  P  K  C  N  L  L  Y  S  K  D  T  E  F  G  Y  Q

    CCAAAACCATCAAGAATGACATGCTGTGCGCCGGCTTCGAGGAGGGCAAGAAGGATGCCT 600
181 P  K  T  I  K  N  D  M  L  C  A  G  F  E  E  G  K  K  D  A

    GCAAGGGCGACTCGGGCGGCCCCCTGGTGTGCCTCGTGGGTCAGTCGTGGCTGCAGGCGG 660
201 C  K  G  D  S  G  G  P  L  V  C  L  V  G  Q  S  W  L  Q  A

    GGGTGATCAGCTGGGGTGAGGGCTGTGCCCGCCAGAACCGCCCAGGTGTCTACATCCGTG 720
221 G  V  I  S  W  G  E  G  C  A  R  Q  N  R  P  G  V  Y  I  R

    TCACCGCCCACCACAACTGGATCCATCGGATCATCCCCAAACTGCAGTTCCAGCCAGCGA 780
241 V  T  A  H  H  N  W  I  H  R  I  I  P  K  L  Q  F  Q  P  A

    GGTTGGGCGGCCAGAAGTGAGACCCCCGGGGCCAGGAGCCCCTTGAGCAGAGCTCTGCAC 840
261 R  L  G  G  Q  K  *  D  P  R  G  Q  E  P  L  E  Q  S  S  A

    CCAGCCTGCCCGCCCACACCATCCTGCTGGTCCTCCCAGCGCTGCTGTTGCACCTGTGAG 900
281 P  S  L  P  A  H  T  I  L  L  V  L  P  A  L  L  L  H  L

    CCCCACCAGACTCATTTGTAAATAGCGCTCCTTCCTCCCCTCTCAAATACCCTTATTTTA 960
    TTTATGTTTCTCCCAATAAA
```

FIG 20C(B)


Hydrophilicity Window Size = 7 Scale = Kyte-Doolittle
50
100
150
200
250
5.00
4.00
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
Hydrophilicity



Luciferase Positive Control
BCON3 EST
221
133
93
67
56
42
BCON3

FIG 21

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 98/00085

A.	CLASSIFICATION OF SUBJECT MATTER
Int Cl^{6}:	C12N 009/12, 009/64, 015/54, 015/57; C07K 016/40; A61K 038/45, 038/48; C12Q 001/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
STN (DGENE) (see below)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
SEQUENCE DATABASES (see below) MEDLINE (see below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
(as online, STN (DGENE): TGGG[AC] [AGT] [GC] T [AGT] AC [AG] GC [AGT] GC [AGT] CA [CT] TG AND GG [AGT] CA [CT] [AT] [GC] [ACT] GG [ACT] CC [ACT] [CT] T and SWISSPROT, GENBANK, EMBL, PIR: SEQ ID Nos: 6, 8 and 10
MEDLINE: 16p13.3 AND "serine protease"

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Proc. Natl. Acad. Sci. USA 87, pages 960-3 (1990) Hanson, R.D. et al. "A cluster of hematopoetic serine protease genes is found on the same chromosomal band as the human α/S T-cell receptor locus." See whole document, especially page 961 column 2-962 column 1.	1,4,7,10,13,16,19,26, 27,31,32,35,38,46,49, 52,56,57,60,63

[X] Further documents are listed in the continuation of Box C

[ ] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 March 1998	03 APR 1998

Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929	JIM CHAN Telephone No.: (02) 6283 2340

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 98/00085

Box 1 **Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 19, 20, 26, 31, 44
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

   the breadth of the claims was such that it was uneconomical to conduct a search that encompassed the full scope of the claims.

3. ☐ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II **Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Proc. Natl. Acad. Sci. USA 87, pages 3811-5 (1990) Vanderslice, P. et al. "Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine protease family." See whole document	1-4,5-7,10-13,14-16,26-28,31,32-38,44,46-52,55-59,60-63
X	J. Reprod. Fertil. 100, pages 567-75 (1994) Bermudez, D. et al. "Proacrosin as a marker of meiotic and post-meiotic germ cell differentiation: quantitative assessment of human spermatogenesis with a monoclonal antibody." See whole document, in particular Introduction.	46-48, 50, 51, 55
X	J. Biol. Chem. 269(29) pages 18843-8 (1994) Yu, J.X. et al. "Prostasin is a novel human serine protease from seminal fluid." See whole document, in particular discussion.	1-3,5,6,10-12,14,15,27,28,31-34,37,44,46-48,50,51,55-59,61,62
X	J. Biol. Chem. 269 (31) pages 19976-82 (1994) Matsushima, M. et al. "Structural characterisation of porcine enteropeptidase." See whole document, especially figure 4.	1-3,5,6,10-12,14,15,27,28,31-34,37,44,46-48,50,51,55-59,61,62
X	J. Biol. Chem. 270 (22) pages 13483-89 (1995) Yu, J.X. et al. "Molecular cloning, tissue-specific expression, and cellular localisation of human prostasin mRNA." See whole document, especially figure 2, Introduction and Discussion.	1-3,5,6,10-12,14,15,27,28,31-34,36,37,41-43,44,46-48,50,51,55-59,61,62
X	Mol. Reprod. Dev. 43, pages 236-47 (1996) O'Brien, D.A. et al. "Boar proacrosin" expressed in spermatids of transgenic mice does not reach the acrosome and disrupts spermatogenesis." See whole document.	1-3,5,6,10-12,14,15,27,28,31-34,36,37,41-43,44,46-48,50,51,55-59,61,62